A STEREOELECTRONIC AND THERMODYNAMIC STUDY OF β-D-METHYL GLUCOSE CONFORMATIONAL CHANGES RELATED TO ANOMERIC CENTRE REACTIVITY

A dissertation submitted in accordance with the requirements of the

UNIVERSITY OF CAPE TOWN

In the fulfillment of the requirements for the degree of

MASTER OF SCIENCE

By

Sabena Shaik Yusuf

Supervisor: Professor Kevin J. Naidoo

Co-Supervisor: Dr Christopher B. Barnett

February, 2014

A Stereoelectronic and Thermodynamic Study of β-D-Methyl Glucose Conformational Change related to Anomeric Centre Reactivity

Sabena Shaik Yusuf February 2014

ABSTRACT

The major part of this thesis focuses on investigating the rationale for ring deformation of β-D-methyl glucose in glycosidase reactions (for example, cellulose hydrolysis). The investigation is computational and is done in isolation from the enzyme binding pocket and incoming nucleophile. *What is the effect of the C1-O1 bond breaking process on key glucose properties* is the central question asked and answered in this thesis. A battery of *ab initio* methods is used to uncover details of the glucose ring pucker free energy volumes. The free energy volumes were computed using the Free Energy from Adaptive Reaction Coordinate Forces (FEARCF) method. The bond stretch of the C1-O1 bond in β-D-methyl glucose serves as a sugar model for hydrolysis, following the $D_N{*}A_N$ mechanism. The FEARCF method has been employed as it was previously shown to generate molecular sampling traversing all of pucker phase space resulting in a multidimensional free energy surfaces (or volumes). Density functional theory and post SCF analysis have been used to investigate the stereoelectronic changes that occur during ring deformation. In particular, changes involving the anomeric carbon, that is the C1-O1, C1-O5 bond distances, electron densities and charges of the C1, O5 and O1 atoms.

Previously the ring deformation in β-1-4 cellulose hydrolysis reaction was investigated in the scientific computing research unit laboratories and the mechanistic pathway of $^4C_1 \rightarrow {}^4H_5 \rightarrow {}^4E \rightarrow {}^4H_3 \rightarrow E_3$ was proven. It can be postulated that at the initial stage of the reaction, the -1 glucopyranose ring adopts the $^4H_5$ conformer. At the transition state the glucopyranose ring changes into the $^4H_3$ conformer. When ring deformation occurs in this mechanism the torsion angles C5-O5-C1-C2 and C5-O5-C1-H1 are planar. The charge on the anomeric carbon is monitored as the pyranose ring deforms away from the chair conformer. Here it is found that when a conformational change occurs the positive charge at C1 increases, making it more susceptible towards nucleophilic attack. The process of ring deformation increases the stabilization energy of the molecule by increasing the stabilizing energy of the O5 lone-pair donation into the σ* anti-bonding orbital on the C1-O1 bond.

"To my mum

for all the inspiration,

guidance and love from above"

ACKNOWLEDGEMENTS

I would like to express my special appreciation and thanks to my supervisors Prof K. J. Naidoo and Dr C. B. Barnett for their endless support, guidance and research proficiency. All of my lab mates for enduring my countless questions especially Krishna, Ian, Riedaa, Pegah, and Werner who have always provided their unique insight.

A special thanks to my Ma and Gran for their words of encouragement, guidance and love, you are the two greatest women in my life. To my Bhai, Shoby, Dino, Eug, Mama and Mami for their support and numerous inspiring conversations. My niece, nephews and cousins, whom are my ultimate source of laughter and happiness.

I would also like to say thank you to Leo, Kersh, Vela and Sanaya for being there when I needed a confidant. To Leo especially, for the immense understanding, support and love during these past two years.

Thank you to NRF and UCT for providing financial support.

LIST OF ABBREVIATIONS

QM/MM	Quantum Mechanics/Molecular Mechanics
CHARMM	Chemistry at Harvard Macromolecular Mechanics
Å	Angstrom (10^{-10} m)
ns	Nanosecond (10^{-9} s)
NMR	Nuclear Magnetic Resonance
ORD	Optical Rotatory Dispersion
HPLC	High Performance Liquid Chromatography
SCC-DFTB	Self Consistent Charge-Density Functional Tight-Binding
NBO	Natural Bond Orbitals
QTAIM	Quantum Theory of Atoms in Molecules
IUPAC	International Union of Pure and Applied Chemistry
HBON	Hydrogen Bonding
DFT	Density Functional Theory
B3LYP	Becke, three-parameter, Lee-Yang-Parr
HOMO	Highest Occupied Molecular Orbital
LUMO	Lowest Unoccupied Molecular Orbital
VMD	Visual Molecular Dynamics
kcal/mol	Kilocalorie per mole
LP	Lone-Pair
BD*	Anti-Bonding Orbital
T. reesei Cel7A	*Trichoderma reesei* cellobiohydrolase glycoside hydrolase family 7
MM2, MM3	Molecular Mechanics Force Fields
PHLB	Palma-Himmel-Liang-Brady
cc-pVDZ	Correlation consistent- polarized valence double zeta

MP*x*	Moller-Plesset perturbation theory at order *x*
CHELPG	Charges from electrostatic potential using a grid-based method
AMBER	Assisted Model Building with Energy Refinement
GROMOS	Groningen Molecular Simulation

*** Other abbreviations not shown here will be explained in the relevant chapters.**

TABLE OF CONTENTS

INDEX OF FIGURES, TABLES AND SCHEMES

CHAPTER 1

CHAPTER 4

Figure 4.12 Natural Hybrid Orbitals (NHO) of a) C1 (O1), b) O1 (C1), c) C1 (O5) and d) O5 (C1) bonds for β-D-Methyl glucose with the C1-OMe bond length at 1.40 Å and 2.00 Å

Figure 4.13 Conformers that are in the -1 position for glycosidase reactions

CHAPTER 1

INTRODUCTION TO CONFORMATIONAL AND STEREOELECTRONIC PROPERTIES OF MONOSACCHARIDES

1.1 THE ROLE OF SACCHARIDES IN PLANTS AND ANIMALS

Carbohydrates play a life-sustaining role in plants and animals. They are a source of energy, provide cell structure, (for example, in the cell wall) and are important in immune response mechanisms.[1] Carbohydrates can consist out of a single monomer (monosaccharide) or they can be more complex and consist of two or more monosaccharide units are that bonded together via glycosidic linkages to form *di*, *oligo, and poly* saccharides. Monosaccharides are the simplest building blocks of carbohydrates and are composed of carbon, hydrogen and oxygen atoms with the empirical formula $(CH_2O)x$. Carbohydrates are abundant in nature (making up most of the organic matter on earth), chiral and highly functionalized.[2] Each carbon that supports a hydroxyl group has a stereogenic center, giving rise to a number of isomeric forms, making these carbohydrate monomers conformationally complex. This intricacy stems from the possibility of many linkages that can be formed via the peripheral hydroxyls of monosaccharide units when they link to each other via glycosidic bonds to form saccharide chains. As well as the multiple stereochemical configurations available to each chiral carbon and the ample range of ring deformations available to monosaccharide units.[3] The nine common monosaccharide units are all derived from D-glucose either via changing the configuration of an exocyclic hydroxyl, removal of the primary alcohol group, ring inversion, or addition of a chemical functional group (see Figure 1.1).

The monosaccharide glucose is one of the main constituents of products of photosynthesis such as starch and a fuel for cellular respiration.[4] Polysaccharides such as cellulose are structural elements in the cell walls of plants and bacteria, giving the cytoplasm in plants shape and structure. The genetic building blocks ribonucleic acid (RNA) and deoxyribonucleic acid (DNA) also contain saccharides. The carbohydrates ribose and deoxyribose provide a structural framework for RNA

and DNA, while the conformational flexibility of these saccharides allow for the storage and expression of genetic information.[3]

Cancer types such as breast; colon and skin cancer have a slightly different distribution of saccharide chains on the cell membrane than normal cells. *β*-N-acetylneuraminic acid (sialic acid) and its derivatives have in particular been implicated in cancer (see work by Varki).[5,6] Inhibition of enzymes that produce these unusual saccharide chains on the cell surface has been shown to reduce tumor growth.[7] The conformational diversity of monosaccharides used to create polysaccharides and biopolymers[8] is one of the reasons they are used extensively in nature.[3]

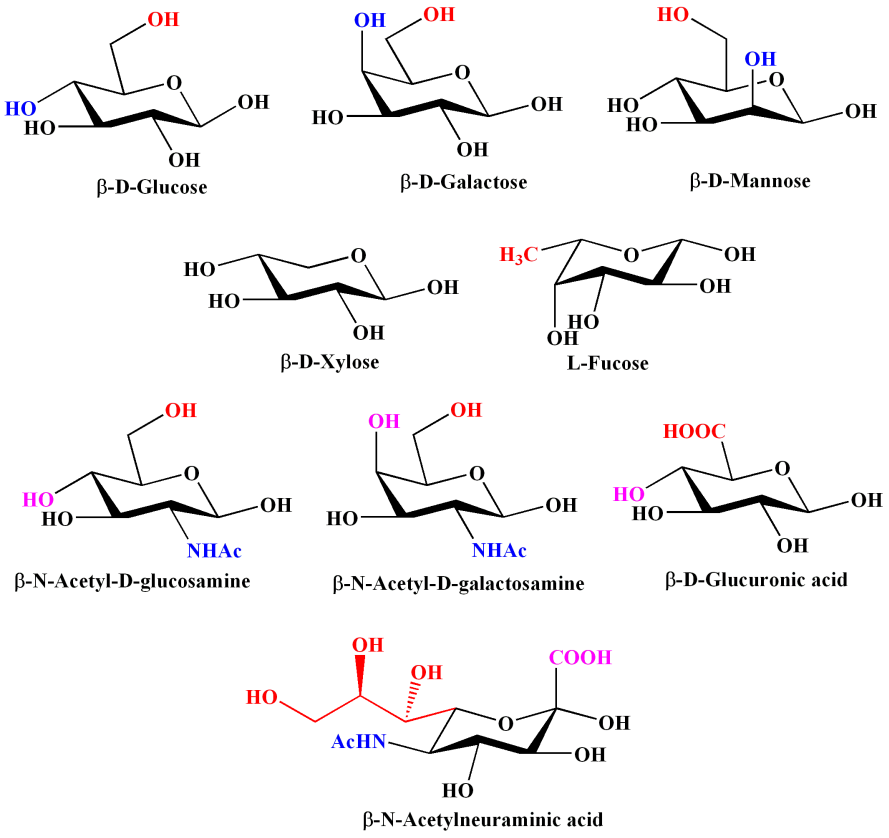


Figure 1.1 Nine common monosaccharide units, with a change in configuration of the hydroxyl at positions C4 and C2 (in blue) for galactose and mannose, the primary alcohol has been replaced with a carboxylic group in glucuronic acid and a hydroxyl chain in *N*-acetylneuraminic acid (shown in red)

1.2 RING CLOSING OF MONOSACCHARIDES

Monosaccharides with 4, 5, and 6 carbon atoms are often called *tetroses*, *pentoses* and *hexoses*. A common hexose sugar is D-glucose, a 6 membered pyranose ring.[9] These sugars have more than one chiral carbon atom. In order to distinguish the different enantiomers, the notation D and L is used to refer to the absolute configuration of the asymmetric carbon furthest from the aldehyde or ketone group in the open chain form.[2,10] In the open chain form there is free rotation about carbon-carbon single bonds. This free rotation can result in close approximation of certain groups in the molecule.[11] Ring formation occurs when the carbonyl group reacts with a hydroxyl group from the other end of the molecule as shown in Figure 1.2. During ring formation the hydroxyl group from the other end of the molecule can attack the carbonyl group via two faces resulting in either the beta or alpha configuration at the anomeric position.[2] The stability of these conformers will be discussed later in this chapter.

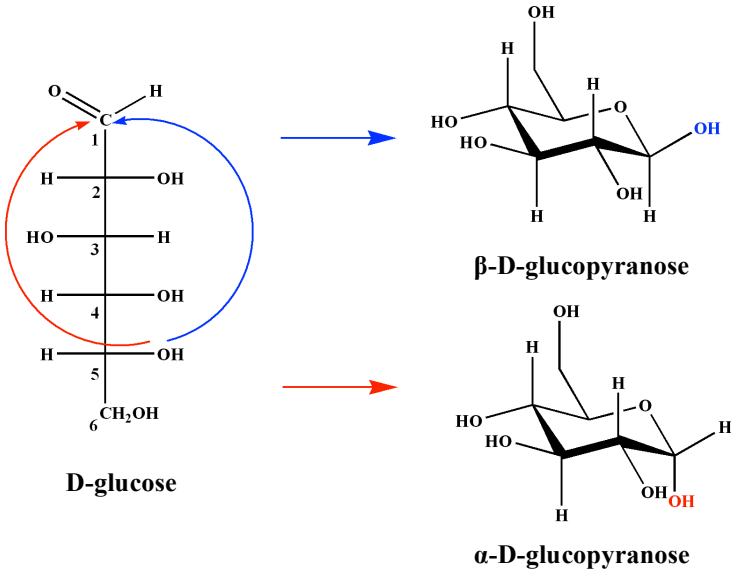
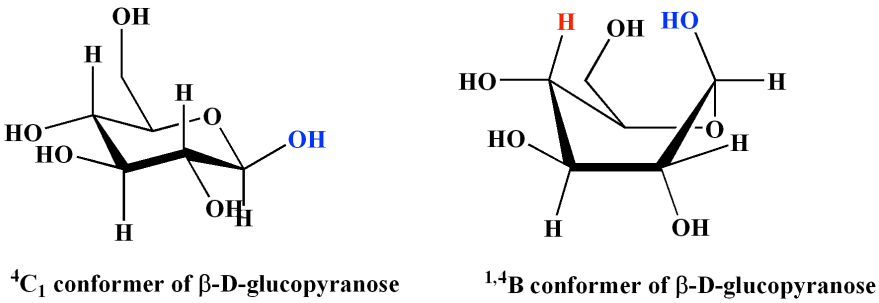


Figure 1.2 Schematic illustration of the formation of the pyranose ring via attack of the hydroxyl group on either face of the carbonyl group (adapted from[12,13]) with the resulting axial (red) and equatorial (blue) positions of the hydroxyl groups at the anomeric position

1.3 CONFORMATION OF PYRANOSE RINGS

Pyranose rings can adopt chair (C), boat (B), twist boat (S), half chair (H) and envelope (E) conformers, with some conformers being higher in energy than others. Together, they constitute 38 ring puckers that depend on the position of atoms defined in the ring. The substituents in the ring can adopt either the axial or equatorial positions. Axial substituents other than hydrogen can cause steric clashes between the orbitals on either axial atom. The chair conformer in Figure 1.3 has the axial positions occupied by the smaller hydrogen atoms rather than, the larger hydroxyl (OH) and primary alcohol groups (CH_2OH) thereby reducing the steric interactions.[2] In the boat conformation the C2, C3, C5 and C6 carbons are coplanar while C1 and C4 are displaced away from the plane of the ring in the same direction as per Figure 1.3. The interactions between the axial hydrogen and hydroxyl group depicted in Figure 1.3 as the red and blue regions, produce a steric strain. There is also torsional strain between the C2-C3 and C5-C6 bonds, which are eclipsed making the boat conformer higher in energy than the chair.[13,14]



$^{4}C_1$ conformer of β-D-glucopyranose $^{1,4}B$ conformer of β-D-glucopyranose

Figure 1.3 A chair and boat conformer of *β*-D-glucopyranose (adapted from[13]), with the change in orientation of the hydrogen atom at C4 (red) and hydroxyl group at C2 (blue)

1.4 SACCHARIDE UNITS IN GLYCOCONJUGATES

Saccharide units present in glycoproteins, glycolipids and other glycoconjugates have distinguishing roles in a variety of molecular processes including, cell-cell interactions, cell adhesion, modulation of growth factor receptors, immune response, inflammation, viral and parasitic infections, as well as in differentiation, development, regulation and many other transduction events. The term glycans can

refer to free sugars or the carbohydrate fragments attached to proteins and lipids (glycoconjugates). Glycans are exposed on the surface of biomolecules and cells and consist of O or N glycosidic linkages of monosaccharide units, which can form large, flexible, branched structures with molecular weights of up to 3 kDa each. The O or N linked glycans are covalently bonded to polypeptide side chains of glycoproteins. N-glycans can be modified without appreciable effects on the protein, allowing for extensive modification and fine-tuning of the biophysical and biological properties of glycoproteins.[8] This extensive modification of glycans allows for cell types and cells in different stages of differentiation and transformation to imprint on their glycoprotein pool their own specific biochemical characteristic.[15]

1.5 SYNTHESIZING GLYCANS

Glycosyltransferases are the enzymes responsible for the formation of glycosidic bonds in carbohydrates. The enzymatic reaction involves the transfer of a monosaccharide unit from an activated donor sugar to a saccharide, protein, lipid or nucleic acid. Donor sugars substrates are most commonly activated in the form of nucleotide diphosphate sugar, however monophosphate, lipid phosphates and unsubstituted phosphates are also used.[16,17] Glycosyl transfer most frequently occurs to the nucleophilic oxygen of a hydroxyl substituent of the acceptor, can also occur to nitrogen, carbon or sulfur nucleophiles.[17] Galactosyltransferases are of biological and medicinal interest as they are involved in the biosynthesis of many cell surface oligosaccharides structures such as blood group antigens and sialyl Lewis X antigens.[18] Increase or decrease of β-1-4-galactosyltransferase (β-1-4-GalT) activity has been associated with disease states such as arthritis and cancer.[17]

1.6 HYDROLYZING GLYCANS

Glycosyl hydrolase (Glycosidases) are the enzymes that break down glycans or the saccharide portion of glycoconjugates. These enzymes catalyse the degradation of carbohydrates that vertebrates ingest. The hydrolysis of carbohydrates releases vital monosaccharide units such as glucose. Although glycosidases can be classified into 100 families, they all use similar mechanisms to catalyse the hydrolysis of glycosidic linkages. Classification of glycosidases according to their specific features of sequence, three-dimensional structure, substrate specificity and reaction mechanism is required as a predictive model for newly discovered enzymes.[19] The simplest classification of glycosidases is according to their substrate specificities.[19] For example β-glucosidase and β-galactosidase are differentiated according to their substrate preference for either a β-D-glucose or a β-D-galactose.[19,20] The drawback of this type of classification is that it does not appropriately accommodate enzymes, which act on several substrates. There are also many structurally unrelated enzymes that have similar substrate specificity. Henrissat et al. provided a detailed analysis of glycoside hydrolases with respect to their catalytic mechanism, mode of action (describes whether the enzyme hydrolyses one of the terminal saccharide units or somewhere within the polymer chain) and amino acid sequence similarities.[19,21,22] The amino acid classification was utilized since there is useful structural and mechanistic information that can be can be derived from amino acid sequences.[22]

The breakage of a glycosidic bond occurs via general acid catalysis with the use of proton donor (acid) and nucleophile (base).[23,24] There are two key mechanisms which result in either the retention or inversion of the configuration at the anomeric carbon.[24] In the enzymatic reaction the glutamic and aspartic amino acids have been realized to play the roles of the catalytic acid or base.[15,25]

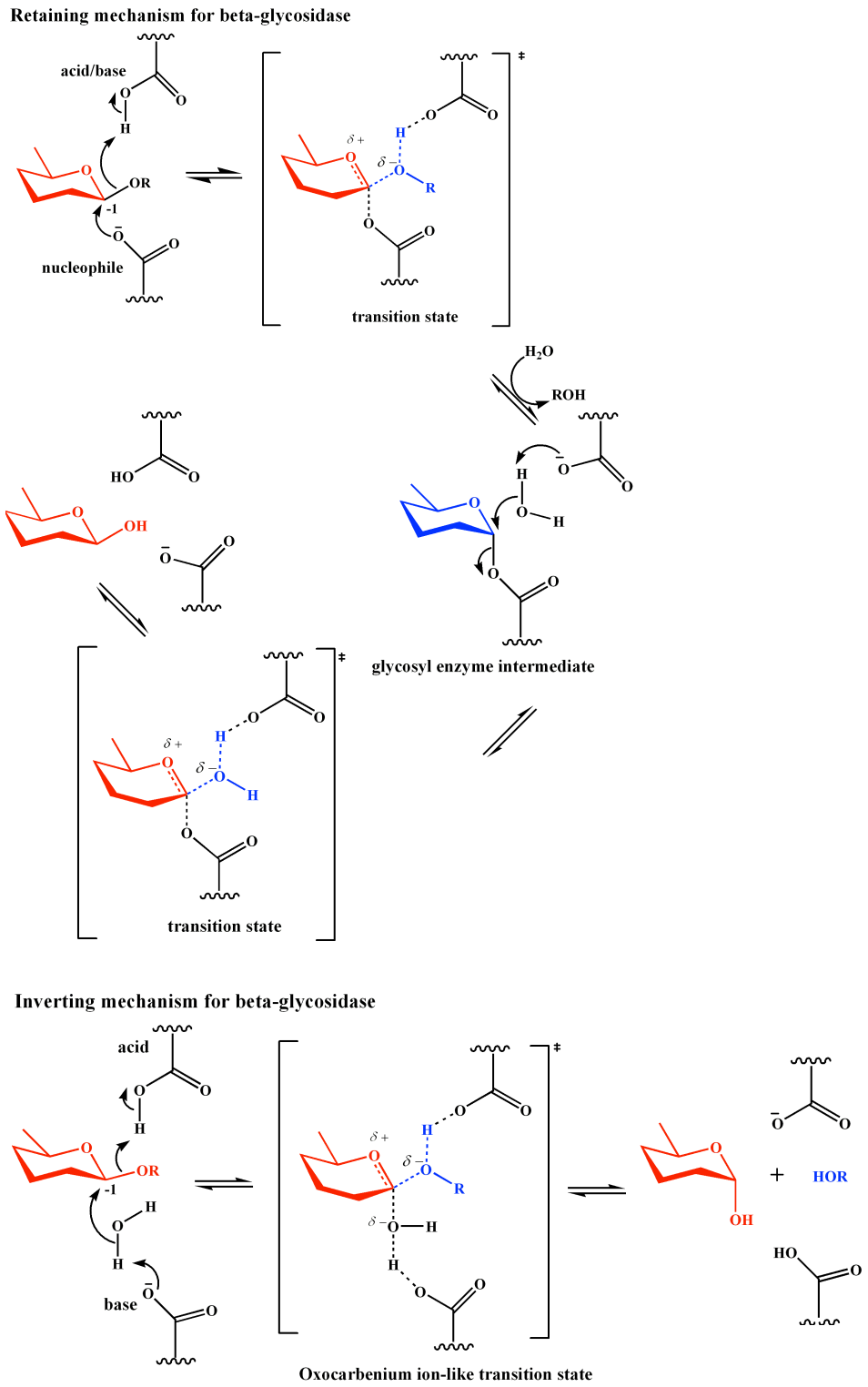


Figure 1.4 Glycosyl hydrolase postulated reaction mechanism for inversion and retention of the configuration at the anomeric carbon (adapted from[23]), with the attack of the nucleophile at the -1 position (reactive monosaccharide, simplified representation since the exocyclic hydroxyls and primary alcohol are not shown) on the saccharide chain. The R group attached at O1 is a mono- or polysaccharide unit. The monosaccharide unit in the transition states (oxo-carbenium ion represented in red) is puckered.

The inverting glycosidase utilizes a direct displacement S_N2 like reaction mechanism that results in an inverted anomeric carbon configuration via an oxo-carbenium ion transition state (shown in red in Figure 1.4), the retaining glycosidase uses a double displacement covalent enzyme intermediate reaction mechanism that results in the retention of the anomeric configuration.

There is a single transition state for the inverting glycosidase reaction and a two-step reaction with a glycosyl intermediate for the retaining glycosidase reaction, both which are thought to possess substantial oxo-carbenium ion-like character, proposed by D. E. Koshland 1953.[17] Davies and co-workers[26] suggest that there is de-localization of lone-pair electrons from the ring oxygen, which stabilizes the postulated cationic transition state. They propose that at or close to, the transition state the pyranose ring will be distorted away from its lowest energy conformation to one that favors such orbital overlap. The ring oxygen's axial lone-pairs donate electrons into the σ* anti-bonding orbital of axial C1-O1 bond, thereby stabilizing the oxo-carbenium ion transition state (see Figure 1.5).

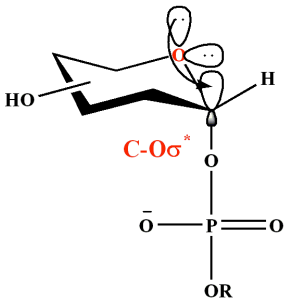


Figure 1.5 The electron donation from the ring oxygen into σ* anti-bonding orbital of the axial C1-O1 bond, phosphorylated at C1, with an axial leaving group (adapted from[27])

This donation of electrons results in a partial double bond character at the C1-O5 bond, which suggests that at or extremely close to the transition state the C2-C1-O5-C5 atoms, must be co-planar. If the substrate monosaccharide does not have the C1-O1 bond axial at the beginning of the reaction, it is postulated that the monosaccharide will be puckered into a conformer that favors electron donation from the ring oxygen into the σ* anti-bonding of the C1-O1 bond shown in Figure 1.5.[26] Such conformations that have the C2-C1-O5-C5 bonds co-planar are seen only for pyranosides in the ^{4}E, E_4, $^{4}H_3$, $^{3}H_4$, $^{2,5}B$ or $B_{2,5}$ conformation.[26]

1.7 CLASSIFICATION OF REACTION MECHANISMS

During nucleophilic displacement that involves the attack of an electron deficient electrophile by an electron rich nucleophile and the removal of the leaving group, the reaction can proceed either via an S_N1 or S_N2 mechanism or via a mechanism that describes a mixture of both of these mechanisms.[28] S_N represents nucleophilic substitution and the notation 1 and 2 are the number of the molecules involved in the rate-determining step.[29] The S_N1 mechanism describes the departure of the leaving group (before the attack of the nucleophile) with the formation of an oxo-carbenium ion like transition state, which is unstable, and is followed by the concomitant attack of the electrophile by the nucleophile leading to the products.[29] The S_N2 mechanism however involves the simultaneous attack of the nucleophile (bond-formation) and departure of the leaving group (bond breakage); therefore the rate-determining step is bimolecular since the rate of the reaction depends on concentration of both the nucleophile and the substrate. This mechanism results in a pentacoordinate slightly sp^2 hybridized anomeric carbon transition state.[29] The nucleophile attacks the electrophilic carbon at an angle of 180^o to the leaving group, since this orientation provides the best overlap between the nucleophile's lone-pair and the σ* antibonding orbital on the leaving group. In glycosidase reactions there is sometimes character of each S_N mechanism in the reaction mechanism which were observed in purine nucleoside phosphorylase[30,31] and purine/pyrimidine phosphoribosyltransferases reactions.[32-35] This analogy has lead to the A_ND_N notation, where the D and A denotes disassociate and associate mechanisms respectively. The A_ND_N mechanism considers the approach of the nucleophile towards the electrophile with the simultaneous departure of the leaving group.[28] If the nucleophile approaches the electrophile and then the slow departure of the leaving group occurs, it's considered to follow an A_N*D_N mechanism (see Figure 1.6).

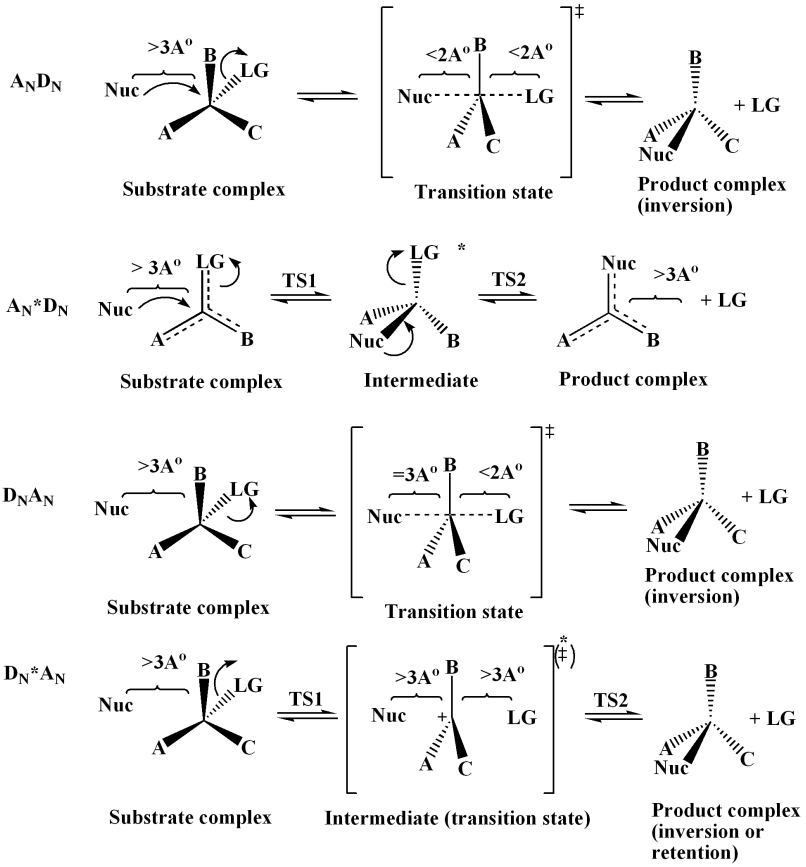


Figure 1.6 The geometric characterization of the atomic motion of the nucleophile and leaving group in enzymatic nucleophilic displacements and substitution reactions using the IUPAC nomenclature (inserted from[28])

The departure of the leaving group can occur before the approach of the nucleophile ($D_N{}^*A_N$), however if the nucleophile attacks before the departure of the leaving group is complete then the mechanism is considered to follow the D_NA_N mechanism. For the $D_N{}^*A_N$ mechanism a transition state is represented by * and an intermediate by ‡. A transition state or intermediate can occur depending on the lifetime of each. An intermediate has a longer lifespan than a transition state, it is also a local minimum on the reaction surface allowing it to be experimentally isolated.[28] Each of the mechanisms described are geometrically characterized by the transition states or intermediate with different distances of the nucleophile and leaving group.[28] In glycosidase reactions the leaving group is the OR (R=monosaccharide or glycan) bonded to the reactive monosaccharide.

Davies et al. and Flint et al. have shown that glycosyl hydrolase enzymes are adapted to utilize high-energy conformations before hydrolysis such as $B_{3,O}$, $^{1}S_{3}$, $^{1}S_{5}$ and $^{1,4}B$.[36-38] In many glycosyl hydrolase retaining and inverting complexes, the substrate was found to adopt a distorted boat or twist boat conformer rather than the low energy chair conformer.[39,40] It has been shown that this distortion of the ring aids in the hydrolysis of glycosidic bonds.[26] This ring distortion places the glycosidic oxygen in close proximity to the acid/base residues in the enzyme.[39] The glycosidic oxygen is the oxygen involved in the glycosidic linkage. It also reduces the steric interaction between the hydrogen on C1 and the incoming nucleophile. In addition it also places the leaving group in a position that facilitates nucleophilic attack. The ring also prefers to distort into a $^{1}S_{3}$ conformer since a conformational change from a $^{4}C_{1} \rightarrow {}^{4}H_{3} \rightarrow {}^{1}S_{3}$ results in the use of the $^{4}H_{3}$ half chair for some monosaccharide units. The $^{4}H_{3}$ conformer places the C5-O5-C1-C2 atoms co-planar, this allows for electron donation and stabilization of the oxo-carbenium ion transition state. The $^{1}S_{3}$ conformer also removes potential steric hindrance between H1 and H3 which are 1,3 diaxial interactions.[26]

1.8 CONFORMATIONAL SPACE AND STEREOELECTRONIC EFFECTS OF MONOSACCHARIDES

Monosaccharides can adopt many different conformations, which have different energy barriers depending on their conformational space. There are stereoelectronic effects that stabilize the ring as well as effects based through space interactions with the environment; for example hydrogen bonding between hydroxyl groups, with solvent or with amino acids in an enzymatic binding pocket. The stereoelectronic and conformational effects also stabilize transition states that are involved in the mechanism of glycosidic bond formation and hydrolysis. Particular configurational and conformational preferences can only be explained in terms of stereoelectronic effects and in this section the anomeric effect, exo-anomeric effect and the gauche effect will be discussed.

1.8.1 ENVIRONMENTAL EFFECTS

The complex electro-structural properties of carbohydrates can change depending on what surrounds the molecule. For instance solvent effects, enzymatic active sites and pH change can induce stereo-electronic conformational change in these molecules. The main aim of this thesis is to understand the behavior of sugar hydrolysis as result of implicit chemical character of a prototypical sugar entity. Nevertheless it is important to appreciate the complexity of sugars resulting from the surrounding environment and thus it is discussed here. Most saccharide units are present in aqueous surroundings. The interactions between monosaccharide units and the water molecules have a direct effect on the structural transformations of carbohydrates and have been studied extensively in water.[41-45] Sometimes explicit water models are used and sometimes implicit water models. Explicit water models use discrete solvent molecules[46] whilst the implicit water model uses a continuous medium, which has the averaged properties of the real solvent.[47] These authors[41-45] addressed some conformational issues that arise when studying carbohydrates in solution, the primary alcohol rotation, structuring of the water molecules around the pyranose ring,[42] relative free energies of the three staggered conformers of the primary alcohol group[45] and the glycosidic bond rotation.[43]

There are many interactions that are at play, and they arise mainly from the multiple hydroxyl and sometimes other functional groups that are found on the periphery of the cyclic monosaccharides. These include electrostatic and van der Waals interactions.[48] Hydrogen bonding interactions can also affect conformational preferences in pyranose rings. Hydrogen bonding is an especially important interaction in protic solvents such as water and in the binding pocket of the enzymes where amino acids that can act as hydrogen bonding donors and acceptors are abundant. For instance the monosaccharide ring changes conformation (pucker) as a result of the hydrogen bonding and van der Waals interactions with the amino acids of the enzyme and the substrate.

1.8.2 THE ANOMERIC EFFECT

In most glucose derivatives one expects a substituent to adopt the equatorial position rather than the axial position. Since there are unfavorable steric interactions between the axial substituent and other functional groups on the ring causing it to be energetically disfavored when compared to the equatorial position.[49] The chair conformer with the largest number of equatorial substituents as shown in Figure1.3, should be favored since there are minimal gauche interactions between substituents. This does occur for a pyranose ring with one hydroxyl substituent, where the equatorial chair conformer is favored. In pyranose rings with multiple hydroxyl groups, this is not always the case. For example, in D-glucose with the hydroxyl group at C1, the equatorial (β) form is favored at equilibrium, when a bulkier substituent is attached to the anomeric carbon such as a methoxy group, the equatorial configuration is no longer favored (as expected) and the axial (α) anomer becomes more dominant.[50] Other electronegative substituents attached to C1 such as chlorine and bromine also favor the axial position. This preference of the electronegative substituent for the axial configuration at the anomeric carbon is termed the anomeric effect.

The anomeric effect occurs in pyranose rings because there is a low-lying σ* anti-bonding orbital that the ring oxygen lone-pairs can donate into. This donation of lone-pair electrons into low-lying anti-bonding orbital of the C1-OR bond can only occur if the substituent at C1 is axial.[50] Since lone-pairs that are orientated equatorially to the ring oxygen are parallel with only the bonds in the ring, the lone-pairs orientated in an axial position are the only electrons that can donate into the antibonding orbital (see Figure 1.5). There is a stabilizing interaction when the donation of electrons from the ring oxygen into the σ* anti-bonding orbital of the C1-OR bond occurs.

This effect possibly occurs during the reaction mechanisms of glycosyltransferase and glycosyl hydrolase enzymes that utilize the oxo-carbenium ion transition state.[14] The *β*-D-glucose unit in the $^4C_1$ conformer does not have the C1-OR bond axial and will probably have to change conformation so that the bond is placed axial in order for the anomeric effect to occur. If the anomeric effect occurs in pyranose rings the

lone-pair on the O5 atom will have to donate electrons into the σ* anti-bonding orbital of the C1-O1 bond with an increased stabilization energy and hence an increase in the stabilization energy of the system.

1.8.3 EXO-ANOMERIC EFFECT

Like the anomeric effect, the exo-anomeric effect displays a preference for the configuration of the substituent at the anomeric carbon. In pyranose rings, the exo-anomeric effect influences the orientation of the R group at C1 in respect to the C1-O5 bond.[51] There is delocalization of lone-pair density on the exocyclic oxygen (O1) atom at C1 into the σ* anti-bonding orbital of the C1-O5 bond.[52] The donation of lone-pair electrons from the exocyclic oxygen at C1 occurs best when the O1-R bond is gauche to the C1-O5 bond, in other words the R group is gauche to O5, see Figure 1.7 (*gg*, *gt*). There is a stabilizing interaction when the donation of electrons from the exocyclic oxygen into the σ* anti-bonding orbital of the C1-O5 bond occurs in the gauche orientation of the R group.

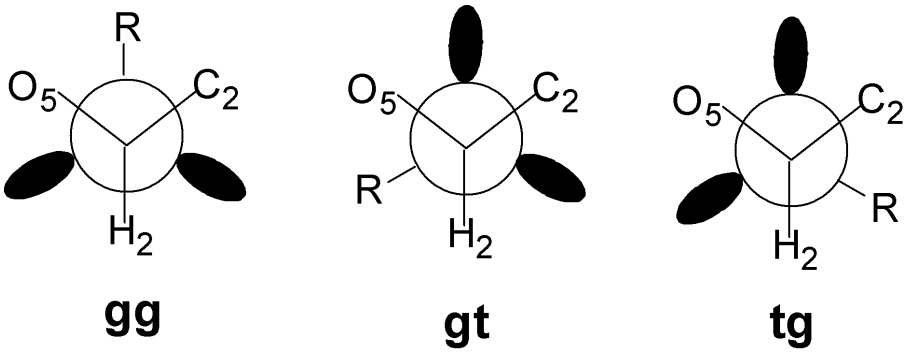


Figure 1.7 Newman projection representation at C1 showing the orientation of R relative to O5 with the lone-pairs on O1 represented as black discs. From left *gg*, *gt*, *tg* (adapted from[51])

There are three staggered conformations about the rotation of the glycosidic bond in both axial and equatorial orientations of the R group. The anomeric carbon can be positioned axial or equatorial and the R group can be orientated *trans-gauche* (*tg*), *gauche-gauche* (*gg*) and *gauche-trans* (*gt*) in respect to the C1-O5 bond. Since the exo-anomeric effect results in a preference for the gauche conformation of the R group, the conformers that place the R group gauche and minimize steric clashes are

expected to be predominant. The conformers that position the R group gauche to the O5 with minimal steric interaction are AGT (axial *gauche-trans*) and EGT (equatorial *gauche trans*).[51] The exo-anomeric effect increases the bond distance of the C1-O5 and shortens the C1-OR bond, since donation of electrons occur from O1 atom into the σ* anti-bonding orbital of the C1-O5 bond, which results in an increase in electron density across the C1-OR bond.

1.8.4 THE GAUCHE EFFECT - PRIMARY ALCOHOL ORIENTATION

The orientation of the hydroxyl methyl (primary alcohol) group in relation to the pyranose ring is defined by two dihedral angles, O5-C5-C6-O6 (ω) and C4-C5-C6-O6 (ω2) (see Figure 1.8). It is more flexible than the other hydroxyl groups since it has an extra carbon atom, which allows for more rotational freedom. The primary alcohol group can adopt three staggered conformers about the ω-angle in relation to the ω2-angle, *gg* (-60^{o}), *gt* (+60^{o}) and *tg* (+180^{o}).[53,54] The gauche effect in monosaccharides is the preference to adopt gauche conformations about the dihedral angle ω or to maximize the number of gauche interactions between electron pairs and/or polar bonds.[53] A gauche interaction occurs when the primary alcohol group orientates itself so that the best donor lone-pair or bond is placed anti-periplanar to the best acceptor bond, which is a stereoelectronic preference. For example, in the *gg* conformer the C6-H6 bond (a good bonding donor) is positioned anti-periplanar to the C5-O5 σ* orbital (a good antibonding acceptor) and the C5-H5 bond (a good bonding donor) is anti-periplanar to the C6-O6 σ* orbital (a good antibonding acceptor).[53] There are staggered conformations of the primary alcohol group where the best donor and acceptor bonds are not anti to one another and therefore are not able to donate electrons into the antibonding orbitals. The conformer that allows for the maximum number of gauche interactions will therefore be preferred.

The *gauche-trans* nomenclature also applies to all the exocyclic hydroxyls on the pyranose ring.

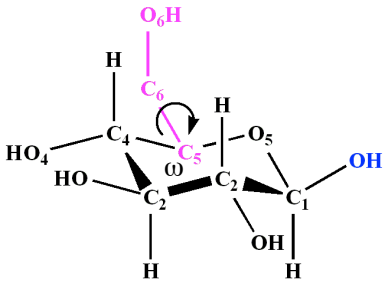


Figure 1.8 A representative of *β*-D-glucose with the primary alcohol torsional angle indicated by ω (adapted from[53])

The torsional angle ω and $\omega 2$ refers to the relation between the O5 and O6 atoms and the C4 and O6 atoms respectively, as given by the dihedral angles O5-C5-C6-O6 and C4-C5-C6-O6. The rotamers are shown in Figure 1.9.

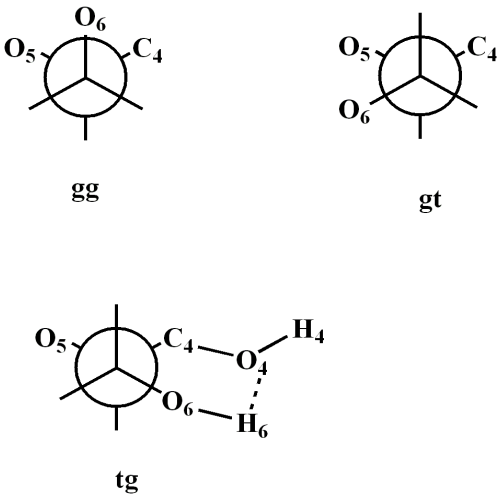


Figure 1.9 Newman projection representation of the primary alcohol rotamers for β-D-glucose. From left *gg, gt, tg*. The dashed lines indicate a potential hydrogen bonding interaction between H6 and O4 in the *tg* rotamer (adapted from[53])

The exocyclic orientations are of interest since different pucker conformations (especially high-energy conformers) may choose a rotamer preference unique to the other conformers. The change in preference could be a result of a potential hydrogen bonding interaction as seen in Figure 1.9. The orientation of each exocyclic hydroxyl and their effect on ring pucker will be considered in this research.

1.9 EXPERIMENTAL TECHNIQUES USED TO STUDY SACCHARIDES

Saccharides are complex, large biomolecules that are difficult to synthesize and manipulate. There are a number techniques used to study carbohydrates experimentally some of which include crystallographic techniques, NMR[55], ORD[56], fluorescence detected HPLC[57] and neutron diffraction.[58] These experimental techniques do have their drawbacks such as difficulty in the structural assignment of complex carbohydrate data by NMR[59], reproduction of the correct conformational ratios that are observed in solution by X-ray crystallography and the crystal induced method used to crystallize polysaccharides can affect the conformation observed in the crystal.[60] However important information can be derived from these techniques. NMR techniques can give insight into conformational aspects of saccharides and structure elucidation by coupling information obtained with mass spectrometry data. ORD techniques are often used to measure the optical rotation with plane polarized light which can determine equilibrium anomeric (α:β) ratios.[61] Fluorescence detected HPLC can provide information on glycosyltransferase/glycosyl hydrolase composition and activity.[62] Neutron diffraction is similar to X-ray diffraction, but in neutron diffraction a beam of thermal or cold neutrons bombard the sample giving a diffraction pattern that provides structural information. This technique is used to study CH···O hydrogen bonds, which plays an important role in crystal packing of carbohydrate structures.[62]

1.10 COMPUTATIONAL AND THEORETICAL METHODS APPLIED TO SACCHARIDES

The computational methods that have been used to study carbohydrates have included Molecular Dynamics (MD), Quantum Mechanics (QM), Molecular Mechanics (MM) and coupled QM/MM techniques (covered in Chapter 2) and Free Energy simulations (covered in Chapter 2 and 3). A successful application of carbohydrate modeling is the degree to which the interatomic interactions can be approximated by a mathematical description. In Molecular Dynamics simulations this is often done using force fields due to the complexity and time requirements of QM methods for large systems. However, with improvements in computing technology such as the use of high performance and parallel computing, it is now becoming more common for

scientists to use QM/MM classical Molecular Dynamics or even QM dynamics – for example Carr-Parinello dynamics.[63] MD and MM have been used to study α-1-4 glycosidic linkages[64] as well as the primary alcohol rotation.[65] A QM approach has been used to reparameterize carbohydrate force fields with a semi-empirical method[66] and the analysis of static QM calculations has been used to investigate the relationship between intramoleculer hydrogen bonding and electron density.[67]

Enzyme based biofuels are a popular focus for solving our global energy demand. Cells can generate energy from an organic molecule such as glucose by oxidizing it completely to CO_2 with the use of an enzyme.[68] Cellulose is a source of glucose units and an abundant biopolymer on earth. Understanding the molecular-level mechanistic process of how enzymes recognize and degrade cellulose can enable the development of engineered enzymes for a cheaper production of glucose, which is used as a biofuel.[69-73] Cellobiohydrolase (cellulase) is the enzyme that hydrolyses cellulose; it consists of carbohydrate binding module (CBM) and a catalytic domain.[73] Experimental techniques such as NMR[74] and X-ray diffraction[75,76] have been used to probe the CBM and catalytic domain. Computational studies have also been used to study the hydrolysis reaction catalyzed by cellobiohydrolase, for example calculation of the potential energy surface of the CBM from *T. reesei* Cel7A,[73] use of coarse-grained models to study the CBM on the hydrophobic cellulose face,[77] and the use of DFT free energy calculations for conformational studies of *T. reesei* Cel7A.[78]

1.11 OBJECTIVES

The primary goal of this thesis is to understand the mechanism for ring deformation as it occurs during glycosidase reactions in particular β-1-4 cellulose hydrolysis. In carbohydrate chemistry a prototypical monosaccharide is glucose. In this thesis a variation in bond length is investigated for the C1-O1 bond of β-D-methyl glucose, taking it as a model for hydrolysis see Figure 1.10.

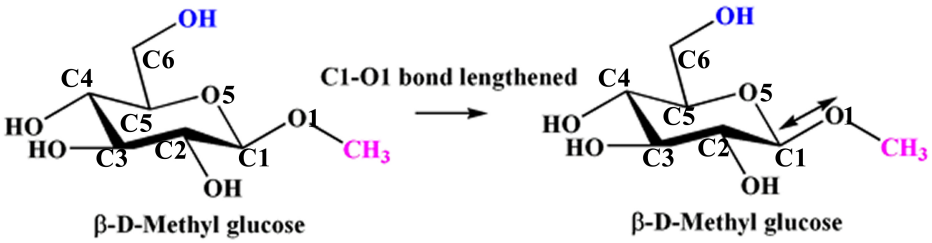


Figure 1.10 Illustration of the C1-O1 bond stretch from 1.40 – 2.40 Å

I intend to investigate the following properties of the prototypical monosaccharide unit, as it relates to the reactivity of the molecule:

- Ring pucker
- Electron density
- Charge
- Anomeric and exo-anomeric effect

In Chapter 2, theoretical conceptions of modeling and simulation methods applied in this dissertation are overviewed, such as Molecular Mechanics (MM), Molecular Dynamics (MD) and Quantum Mechanics (QM). More specifically, Section 2.2 contains a description on how to perform a molecular dynamics simulation, while Section 2.3 discusses quantum theory. This general discussion is then followed by density functional theory and SCC-DFTB, a semi-empirical method.

In Chapter 3, the specialized methods used in computational experiments of carbohydrates are discussed, including the parameterized force fields used when studying the exo-anomeric and anomeric effects and the different methods used for calculating free energy. A histogram based free energy method is implemented to calculate free energies of ring pucker while DFT methods are used to study the electronics.

In Chapter 4 the methods developed in the previous two chapters are applied to the sugar model for hydrolysis. The free energy of conformational change is investigated as a function of the β-D-methyl glucose C1-O1 bond length as the reaction coordinate using the Free Energy from Adaptive Reaction Coordinate Forces (FEARCF) method. The electron density across the C1-O5 bond, electronic structure and charge on C1, O1, O5 as a function of the C1-OMe bond length are investigated. It is shown that a change in conformation of the pyranose ring is relevant in the context of cellulose hydrolysis.

Chapter 5 presents the concluding remarks on ring deformation and its importance in cellulose hydrolysis as well as possible future work.

CHAPTER 2

MODELLING MOLECULAR SYSTEMS

It is imperative to completely understand the synergy between the macroscopic and microscopic properties of chemical compounds in the environment. Since there are many factors at a microscopic level that can contribute to or determine a specific macroscopic property. The microscopic properties of chemical compounds can be modified to affect its macroscopic property, structure or function. The study of these microscopic properties requires investigation on the atomic/molecular level. Experimental results yielding detailed microscopic information on complex compounds in solution are difficult, if not impossible to perform. Computer simulations, which provide explicit information on the microscopic scale, can be a very powerful tool for investigating the macroscopic motions and interactions of molecules and atoms.

A microstate is a specific configuration that a system may occupy during its thermal fluctuation, and collection of microstates provides one with information to calculate a system's macroscopic properties such as temperature or pressure.[79,80] Molecular dynamic (MD) simulations sample microstates within an ensemble. The ensemble averages can be related to the time averages obtained during an MD simulation by applying the ergodic hypothesis.[81] The ergodic hypothesis states that if a system is left to itself long enough, it will eventually pass through all the dynamical states provided that energy is conserved.[81] The observable obtained from an MD simulation can be directly related to the values for experimentally observed variables.

There are two main categories of chemical simulations: quantum mechanical simulations (*ab initio* and semi-empirical) and classical (molecular) mechanics simulations.[82] Molecular Mechanics (MM) and Molecular Dynamics simulation methods are generally referred to as *force field methods*. A force field is a collection of mathematical functions that are parameterized to give the energy of a system in terms of such as bonds, angles, dihedral angles, improper dihedral angles and non-bonded interactions.[83] All force fields are parameterized based on higher levels of

theory.[84] The MM method does not explicitly account for electrons and their interactions or for the electronic structure of the molecule.[83] The forces on the atoms are calculated from an empirical force field not from the method used to generate ensembles i.e. it is possible to run *ab initio* MD.

A measurement of how the atoms and molecules change over time can be described using classical mechanics, i.e. Newton's equations of motion. MD monitors the evolution of a system over a short period of time also known as a *timestep*, by integrating Newton's equations of motion, in order to obtain the new positions of the particles (see equation 2.1). This results in a trajectory, which shows how the coordinates and velocities of the atoms change over time that can be obtained by solving equation 2.1, which states that the motion of the particle of mass m_i along the coordinate y_i caused by a force F_{yi} acting on the particle in that direction can be derived.[82]

$$F_{y_i} = m_i \frac{d^2 y}{dt^2} \tag{2.1}$$

2.1 MOLECULAR MECHANICS

The energy of the system is calculated as a combination of the bonded and non-bonded interactions present in the system. The total energy is the sum of energy from each bond length, angle, dihedral and non-bonded interactions (represent the influence of non-covalent forces) see Figure 2.1.[85]

$$E_{tot} = E_l + E_\theta + E_\omega + E_{nb} \tag{2.2}$$

The typical vibration of a bond is most accurately described by the Morse function but most current force fields utilize a simple harmonic function

$$E_l = \sum k_l \left(l - l_o\right)^2 \tag{2.3}$$

k_l is the stretching force constant which describes the change in bond length and l_o is the equilibrium bond length. The bond stretch is treated similarly to a stretched

spring.[86] At extended bond lengths or when the bond is almost broken, the harmonic function cannot describe the dissociation.

The bond angles are also described by a harmonic function

$$E_\theta = \sum k_\theta \left(\theta - \theta_o\right)^2 \tag{2.4}$$

As before, the k_θ is the force constant and the θ_o is the equilibrium value for the bond angle.[86] The dihedral angles take into account the *gauche-trans*[87] configurations and are described by the equation

$$E_\omega = \sum V_n \left(1 + s\cos n\omega\right) \tag{2.5}$$

where V_n is the rotational barrier height, n the periodicity of rotation and s = 1 for staggered minima and -1 for eclipsed minima.

The non-bonded interactions are limited to pairwise interactions and are distance-dependent. The van der Waals interaction is commonly described by the Lennard-Jones potential $E_{vdw} = \sum \varepsilon \left[\left(\frac{r_m}{r} \right)^{12} - 2\left(\frac{r_m}{r} \right)^6 \right]$ (2.6)

where ε is the well depth and r_m is the minimum energy interaction distance.[85] The electrostatic non-bonded interaction is calculated using partial charges (q) on the atom centres with energy of the electrostatic interaction calculated by Coulomb's law.

$$E_{el} = \sum \frac{q_i q_j}{D r_{ij}} \tag{2.7}$$

The dielectric constant D depends on the solvent, r_{ij} is the distance between the charge.

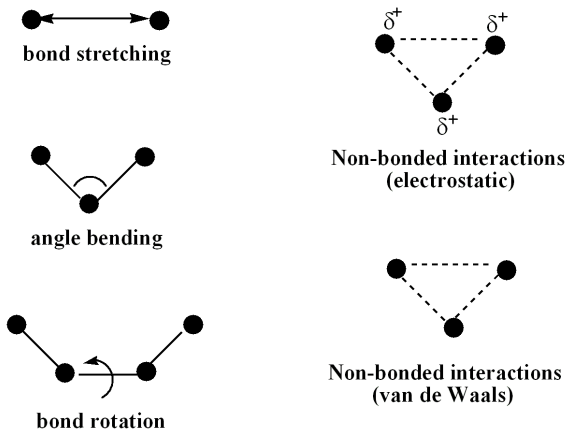


Figure 2.1 Schematic illustration of the different types of bonding and non-bonding interactions: these include bond stretching, angle bending, bond rotation, van der Waals and electrostatic interactions (adapted from[88])

The common additive empirical of force fields have two classes, a potential energy function is added in both classes to describe the relationship of the structure to the energy of the system. The first class of force fields uses an additive potential energy function with a minimal set of forces that can describe molecular structure; bond and angles are treated harmonically and dihedral and torsional rotations are described by a sinusoidal term. Interactions between atoms such as repulsion and dispersion are described by a Lennard-Jones term, with electrostatic interactions treated via a Coulombic term.[89] The terms contributing to the energy function are similar for biomolecular force fields such as CHARMM, AMBER and GROMOS and these are referred to as Class I force fields. Additions or alternate descriptions to the Class I force fields, for bonds, angles and dihedrals terms are referred to as Class II force fields.[90] These additions have increased the accuracy of force fields to treat conformational energies. Some Class II force fields are MM2 and MM3.

To effectively define a force field, it is imperative to express all the parameters shown above in the force field. Molecular mechanics bypasses the solution of the electronic Schrödinger equation and does not account for the quantum aspects of the nuclear motion of electrons and nuclei, although these have been parameterized into the force field to some extent.[82,91]

2.2 MOLECULAR DYNAMICS SIMULATION

Once the force field has been parameterized a molecular dynamics simulation can be performed in order to generate an ensemble. This process involves energy minimization, heating, equilibration and production

The initial energy obtained from the MM calculation maybe high and not representative of the actual structure. The molecular structure is minimized to obtain more reliable geometries and energies. Many of the minimization programs in use today are based on the mathematical principles of the Newton-Raphson method.[85]

During minimization the kinetic energy which is dependent on the velocities of the particle and vital for an MD simulation are not taken into account. Although the potential energy of the system is lowered using energy minimization, the kinetic energy may increases rapidly in an MD simulation that is run immediately after minimization on the potential energy surface leading to unstable dynamics. Therefore atoms need to be gradually heated to room temperature before attempting any MD.

The system is equilibrated, by allowing it to evolve for a period of time and integrating Newton's equations of motion until the systems average temperature and structure is stable. Regular reassigning of velocities that are appropriate to the desired temperature facilitates this and can help to speed up equilibration.

The equilibrated structure is used as the starting point for production dynamics. The simulation is then allowed to evolve spontaneously for a period of time. The hydrogen bonded and non-bonded lists are updated frequently. The hydrogen bonded list is only updated in CHARMM if hydrogen bond corrections are applied to the system.

2.2.1 ENSEMBLES

An ensemble is a subset of all the states sampled by a system. MD simulations can be carried out using constant number of particles (*N*), volume (*V*) and energy (*E*) (referred to as the *micro canonical* or constant *NVE* ensemble). A *canonical* ensemble uses a constant *N*, *V* and *T* (temperature) instead.[92] The *isothermal isobaric*

ensemble can also be used with a constant *N*, *T* and *P* (pressure). In all the three ensembles mentioned above there are a constant number of particles, however in a *grand-canonical* ensemble the composition can change. In this study, all the simulations have been performed using a *canonical* ensemble.

The packages available to perform MD include AMBER[93], CHARMM[94] and GROMOS[95] each with their own applicable force field. Choosing a force field would depend on how accurately the force field under consideration represents the actual system. Force fields are frequently parameterized to a particular subset of molecules and are used to describe the inter-atomic properties of the molecule.[96] The CHARMM program was used for all Molecular Dynamic simulations in this study, with the CHARMM general all-atom force field developed by MacKerell.[97]

2.2.2 CARBOHYDRATE MODELS

The common carbohydrate force fields currently in use, includes those of AMBER[93], CHARMM[94] and GROMOS[95,98], with the specificities discussed in the following chapter. The degree of accuracy to reproduce experimental data of each force field is system dependent. They can perform poorly in molecules that deviate from systems that the force field was parameterized for.[99] The CSFF was developed to improve the primary alcohol rotation to agree more closely to experimental data.[45] This is important when studies are done on the primary alcohol rotation. The CHARMM general all-atom force field were parameterized specifically for carbohydrates, proteins and lipids.[96] The force field parameterized specifically for monosaccharide units was used in this study.[97] Although MM methods have proven useful in obtaining structures and binding energies over the years there are still some shortcomings to this method. Classical mechanics cannot be used to describe even qualitatively correctly the electron distribution since electrons display both wave and particle characteristics.[100] To describe electrons with increased accuracy, quantum calculations are required with a solution to the wave function, Ψ.[82]

2.3 QUANTUM MECHANICS

There are two types of quantum mechanical methods "*ab initio*" and semi-empirical methods. In Latin, *ab initio* means "*from the beginning*". These methods are based upon solutions (obtained from solving the Schrödinger equation) generated from first principles, which does not take into account experimental data, in contrast to semi-empirical methods. Semi-empirical techniques obtain certain parameters from experimental data, with approximations being made to the Hartree-Fock formalism, allowing for these methods to be less time-consuming. *Ab initio* methods are reliable but computationally expensive, and are thus applied to the study of small and medium size systems.[82]

The wave function expressed as the time independent Schrödinger equation is shown in short hand notation in equation 2.8 where $\hat{H}$ is the *Hamiltonian Operator* (a set of operations that allows one to calculate its energy) and E is the energy of the system. The wave function Ψ describes the positions and motions of all nuclei and electrons of the system.[85]

$$\hat{H}\Psi = E\Psi \tag{2.8}$$

for time independent systems we can solve the spatial Schrödinger equation which depends on the coordinates, r.[101] The symbol $\hat{H}$ indicates that the Hamiltonian is an operator that acts on the wave function Ψ. The Hamiltonian is an energy operator that consists of potential and kinetic energy terms.[101-103]

$$\left(\hat{T}+\hat{V}\right)\Psi(r) = E\Psi(r) \tag{2.9}$$

If the potential energy V does not depend on time then we use the time independent Schrödinger equation (equation 2.9). The kinetic energy operator is given as

$$\hat{T} = \frac{-\hbar^2}{2m}\nabla^2 \tag{2.10}$$

where ∇^2 is the Laplacian operator : $\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}$ (2.11)

, while the potential energy term is given by:

$$\hat{V} = \frac{1}{4\pi\varepsilon_0}\left(-\sum_{i}^{electrons}\sum_{i}^{nuclei}\left(\frac{Z_I e^2}{|R_I - r_i|}\right) + \sum_{i}^{electrons}\sum_{j<i}^{electrons}\left(\frac{e^2}{|r_i - r_j|}\right) + \sum_{i}^{nuclei}\sum_{J<I}^{nuclei}\left(\frac{Z_I Z_J e^2}{|R_i - R_j|}\right)\right) \tag{2.12}$$

The potential energy is expressed in terms of e, which is the charge and Z the atomic number. The summations in this term represent electron-nuclei attraction, electron-electron repulsion and nuclear-nuclear repulsion. From equation 2.8, if the Hamiltonian operator is applied to the wave function Ψ, then an energy value is obtained for the molecule, which is referred to as a solution to the Schrödinger equation. The Schrödinger equation can be solved exactly for simple systems such as the hydrogen atom and a particle in a box. However, when applied too more complex systems the solutions to this equation can only be approximated. To solve to the wave equation of larger systems a number of approximations must be made.

2.3.1 THE BORN-OPPENHEIMER APPROXIMATION

The Born-Oppenheimer approximation states that electrons will almost immediately adjust to the nuclear coordinates of a molecule.[104] This assumption is justified because the motions of lighter electrons are much faster than the motion of nuclei. Using the Born-Oppenheimer approximation, the term considering nuclear kinetics in the Hamiltonian can be disregarded. The resultant energy is thus:

$$\hat{H}^e \Psi^e(r,R) = E^{eff}(R)\Psi^e(r,R) \tag{2.13}$$

where R and r are the nuclear positions and electronic coordinates respectively. The electronic Hamiltonian ($\hat{H}^e$) is a sum of kinetic and potential energy terms.

$$\hat{H}^e = \hat{T}^e + \hat{V} \tag{2.14}$$

with the kinetic energy shown in equation 2.15 and the potential energy in equation 2.12

$$\hat{T}^e = -\left(\frac{h^2}{8\pi^2 m}\right)\sum_i^{electrons}\left(\frac{\partial^2}{\partial x_i^2}+\frac{\partial^2}{\partial y_i^2}+\frac{\partial^2}{\partial z_i^2}\right) \tag{2.15}$$

2.3.2 THE ONE-DIMENSIONAL SCHRÖDINGER EQUATION

Consider the Schrödinger equation for a single electron moving in one dimension (along a line), the wave function $\Psi(x)$ is a function of the coordinate x or for every value of x there is a corresponding value of the wave function $\Psi(x)$.[105]

The electron density is obtained by calculating the square of the wave function (see equation 2.16) where $\rho(x)$ is the electron density

$$\rho(x) = |\Psi(x)|^2 \tag{2.16}$$

where

$$|\Psi(x)|^2 = \Psi(x)\Psi^*(x) \tag{2.17}$$

The Schrödinger wave equation for a particle in one dimension is given by equation 2.18.[106]

$$-\frac{\hbar^2}{2m}\frac{d^2\Psi}{dx^2} + V_\Psi = E_\Psi \tag{2.18}$$

where $\hbar$ equals $h/2\pi$ and m is the particle mass, for an electron, $m = 9.109\text{x}10^{-28}$ g. V is the function that gives the potential energy of the electron relative to its position x. E, the energy of the electron is determined together with the wave function (see equation 2.8). The energy term (E) is said to be an eigenvalue. Solutions to the Schrödinger equation obeying certain conditions will only exist for particular values of E. These values of E are the allowed energy levels of the system, which depend on V and m from equation 2.18. Hence, if given the values m, (which particle is being considered e.g. electron, nuclei etc.) and the function V (which specifies the surroundings in which the particle is found), one can calculate the allowed energies. The condition imposed on the wave function is related to the electron density. Ψ must be continuous, since the electron density cannot vary suddenly from point to point. The electron density must have only one value at any particular point. The square of the wave function associated with a particular eigenvalue is the electron probability density for the state pertaining to that energy (equation 2.16).

2.3.3 THE VARIATIONAL PRINCIPLE

The fundamental postulate of quantum mechanics states that a wave function ψ exists for any chemical system and the appropriate operators, which act upon the wave function, will result in the observables of the system. From the molecular wave function it is possible to calculate the energy by applying the appropriate Hamiltonian operator.[107] If it is assumed that we can pick an arbitrary function Φ that has the appropriate nuclear and electronic coordinates to be operated upon by the Hamiltonian. Φ can be formulated as a linear combination of a complete set of orthonormal basis functions, ψ_i as shown in equation 2.19. Orthonormal describes two vectors that are orthogonal. Orthonormal set describes a set of vectors that are orthogonal, if these sets compose a function, then it is termed an orthonormal basis.

$$\Phi = \sum c_i \psi_i \tag{2.19}$$

where the individual ψ_i and c_i are unknown but the orthonormality of Φ puts a limitation on the coefficients thus the energy associated with the generic wave function Φ is determined from all the coefficients c_i and their associated energies E_i (see equation 2.20).[108]

$$\int \Phi H \Phi \, dr = \sum_i c_i^2 E_i \tag{2.20}$$

In the set of associated energies E_i there must be a lowest energy value, which will be referred to as the 'ground state', E_0. According to the Variational Principle, this result must be greater than or equal to zero, thus we have

$$\frac{\int \Phi H \Phi \, dr}{\int \Phi^2 \, dr} \geq E_0 \tag{2.21}$$

The equation 2.21 can be used to judge the quality of the wave functions that are arbitrarily guessed by the associated energies. The lower the associated energy the closer it is to the ground state energy and therefore more accurate.

2.3.4 LINEAR COMBINATION OF ATOMIC ORBITALS (LCAO)

From equation 2.21 it is possible to construct a wave function in any manner that is reasonable and then determine the quality of the wave function in comparison to other wave functions, by evaluating the energy eigenvalues associated with each wave function. The eigenvalue that is the lowest will be the most accurate since it is closest to the ground state energy and presumably the best one to use to obtain other properties of the system by the application of other operators.

For a one electron system, which only has one nucleus, we can solve the equation 2.18 exactly without guessing wave functions. The eigenfunctions that is determined for a one electron system such as the hydrogen atom are the atomic orbitals, 1s, 2s, etc. From this we can construct a guess wave function ϕ as a linear combination of atomic wave functions. $\phi = \sum_{i=1}^{N} a_i \varphi_i$ (2.22)

where the set of N functions φ_i is referred to as a 'basis set' and each has associated with it some coefficient a_i.[109] This assembly is known as the linear combination of atomic orbitals (LCAO).[110] From equation 2.16, the wave function squared gives the probability density, (an estimate of where the electrons are likely to be found). In equation 2.22 the upper limit N is not infinite, however, the more atomic orbitals that are allowed into the basis set the closer we are to estimating the true molecular orbital space.

2.3.5 THE SECULAR EQUATION

The energy of the guess function is as follows,

$$E = \frac{\sum_{ij} a_i a_j H_{ij}}{\sum_{ij} a_i a_j S_{ij}} \qquad (2.23)$$

where the notation H_{ij} and S_{ij} are the integrals in the denominator and numerator, which are the resonance integral and overlap integral, respectively.[107]

The overlap integral is the measure of how much two basis functions overlap in a phase matched way in space. The orbitals that give rise to large overlap will have

large resonance integrals. From the variational principle we can infer that the closer we are to the 'true' one-electron ground state wave function, the lower our energy will be. Once we have selected a basis set we would choose the coefficients a_i so as to minimize the energy for all possible linear combinations of our basis function. We know that for a function to be at its minimum, their derivatives with respect to all its free variables are zero as indicated in equation 2.24.[107]

$$\frac{\partial E}{\partial a_k} = 0 \tag{2.24}$$

Performing the partial differentiation on equation 2.23 for each of the N variables a_k results in N equations which must be met for equation 2.24 to hold true, shown below.

$$\sum_{i=1}^{N} a_i \left(H_{ki} - ES_{ki} \right) = 0 \tag{2.25}$$

These sets of N equations involves N unknowns, it has a non-trivial solution if the determinant formed from the coefficients of the various quantities H_{ki}-ES_{ki} is equal to zero. Numerically, it is shown as

$$\begin{vmatrix} H_{11} - ES_{11} & H_{12} - ES_{12} & \cdots & H_{1N} - ES_{1N} \\ H_{21} - ES_{21} & H_{22} - ES_{22} & \cdots & H_{2N} - ES_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ H_{N1} - ES_{N1} & H_{N2} - ES_{N2} & \cdots & H_{NN} - ES_{NN} \end{vmatrix} = 0 \tag{2.26}$$

which is known as the secular equation; there will be N energies E_j where each value of E_j will give rise to a different set of coefficients a_{ij}, using E_j and this set of coefficients will define an optimal function ϕ_j within the given basis set i.e equation 2.22.

2.3.5 MANY ELECTRON WAVE FUNCTION

From Hückel theory we can derive molecular orbitals and molecular orbital energies using an *one-electron formalism*, we can assume that the energy of a many electron system could be determined by summing the energies of the occupied one-electron orbitals. This theory, however, is rarely sufficient for accurate quantitative

assessments. The Hückel theory assumes that the orbitals themselves are invariant to the number of electrons in the π system. There will however, be repulsion between electrons. This theory partially accounts for electron-electron repulsion in an average way, but it is rather crude.

2.3.5.1 HARTREE-PRODUCT WAVE FUNCTIONS

The Hamiltonian is 'separable' and can be expressed as shown in equation 2.27, where N is the total number of electrons and h_i is the one-electron Hamiltonian defined in equation 2.28.[111]

$$H = \sum_{i=1}^{N} h_i \tag{2.27}$$

$$h_i = -\frac{1}{2}\nabla_i^2 - \sum_{k=1}^{M} \frac{Z_k}{r_{ik}} \tag{2.28}$$

where M is the total number of nuclei, and eigenfunctions of the one-electron Hamiltonian (equation 2.28) must satisfy the corresponding single electron Schrödinger equation

$$h_i \psi_i = \varepsilon_i \psi_i \tag{2.29}$$

As mentioned above the Hamiltonian is separable, therefore its many electron eigenfunctions can be constructed as products of one-electron eigenfunctions, or one electron molecular orbitals shown in equation 2.30.

$$\Psi_0 = \psi_0(1)\psi_0(2)\cdots\psi_0(n) \tag{2.30}$$

This equation is referred to a 'Hartree-product' wave function, where Ψ_0 is a function of all the coordinates of all the electrons in the atom, $\psi_0(1)$ is a function of the coordinates of electron 1, $\psi_0(2)$ is a function of the coordinates of electron 2, etc.; the one electron functions $\psi_0(1)$, $\psi_0(2)$ are atomic orbitals or molecular orbitals if we were analyzing a molecule.

2.3.5.2 HARTREE OPERATOR (HAMILTONIAN)

The Hamiltonian defined earlier in equation 2.28 does not consider the interelectronic repulsion, which depends not on one electron but on all possible pairwise interactions. We would like to find molecular orbitals ψ that minimize $\left\langle \Psi_{HP} \middle| H \middle| \Psi_{HP} \right\rangle$.

Equation 2.31 shows that each molecular orbital ψ_i is an eigenfunction of its own operator h_i defined by

$$h_i = -\frac{1}{2}\nabla_i^2 - \sum_{k=1}^{M} \frac{Z_k}{r_{ik}} + V_i\{j\} \tag{2.31}$$

where the last term represents an interaction potential with all other electrons occupying orbitals j which is defined as

$$V_i\{j\} = \sum_{j \neq i} \int \frac{p_j}{r_{ij}} dr \tag{2.32}$$

where p_j is the electron (probability) density associated with electron j. The repulsive third term in equation 2.31 is analogous to the attractive second term except that nuclei are treated as point charges, while electrons are treated as waves with their charge spread out. Fock and Slater corrected the drawbacks of the Hartree method described above. Hartree's iterative, average field methodology supplemented with electron spin and antisymmetry (described below) lead to the creation of the Hartree-Fock equations.

2.3.6 ELECTRON SPIN AND ANTISYMMETRY

Electrons are characterized by a spin quantum number having the value ± ½. The electron spin function is an eigenfunction of the operator S_z, the spin eigenfunctions are orthonormal and are designated as α and β. From Pauli's exclusion principle, which states that, no two electrons can be characterized by the same set of quantum numbers, in a molecular orbital there are only two choices for the spin quantum number, α or β and thus only two electrons may be placed in any molecular orbital.

As an illustration a ground state Hartree-product wave function having two electrons of the same spin may be constructed, as shown in equation 2.33.[107]

$$^3\Psi_{HP} = \psi_a(1)\alpha(1)\psi_b(2)\alpha(2) \tag{2.33}$$

ψ_a and ψ_b are different from each other and orthonormal. Electronic wave functions must invert the sign whenever the coordinates of two electrons are interchanged, which is referred to as antisymmetry. We can define the permutation operator P_{ij} as the operator that interchanges the coordinates of electrons i and j. Thus we can write the Pauli principle for a system of N electrons shown in equation 2.34

$$\begin{aligned} &P_{ij}\Psi\left[q_1(1),...,q_i(i),...,q_j(j),...,q_N(N)\right] \\ &=\Psi\left[q_1(1),...,q_j(i),...,q_i(j),...,q_N(N)\right] \\ &=-\Psi\left[q_1(1),...,q_i(i),...,q_j(j),...,q_N(N)\right] \end{aligned} \tag{2.34}$$

where q includes the spin function

$$^3\Psi_{SD}=\frac{1}{\sqrt{2}}\left[\psi_a(1)\alpha(1)\psi_b(2)\alpha(2)-\psi_a(2)\alpha(2)\psi_b(1)\alpha(1)\right] \tag{2.35}$$

Equation 2.35 satisfies the Pauli principle. SD stands for 'Slater determinant', which is explained below.[112]

2.3.7 SLATER DETERMINANTS

The Slater determinant ensures that there are no more than two electrons in each spatial orbital. A different mathematical notation is used for equation 2.35, shown below. The difference of molecular orbital products has been expressed as a determinant as shown in equation 2.36

$$^3\Psi_{SD}=\frac{1}{\sqrt{2}}\begin{vmatrix} \psi_a(1)\alpha(1) & \psi_b(1)\alpha(1) \\ \psi_a(2)\alpha(2) & \psi_b(2)\alpha(2) \end{vmatrix} \tag{2.36}$$

The general property of a determinant is that it changes sign when any two rows or columns are interchanged, the use of this feature in constructing antisymmetric wave functions was first discovered by Slater[112]

$$\psi_{SD}=\frac{1}{\sqrt{N!}}\begin{vmatrix} X_1(1) & X_2(1) & \cdots & X_N(1) \\ X_1(2) & X_2(2) & \cdots & X_N(2) \\ \vdots & \vdots & \ddots & \vdots \\ X_1(N) & X_2(N) & \cdots & X_N(N) \end{vmatrix} \tag{2.37}$$

where N is the total number of electrons and X is a spin-orbital, which is a product of a spatial orbital and an electron spin eigenfunction.[112] If two spin orbitals differ only in the spin eigenfunction, it results in a doubly filled orbital. Equation 2.37 has every electron appearing in every spin orbital, this is a result of the indistinguishability of quantum particles. Let's consider the energy of interelectronic repulsion for the wave function described in equation 2.28, evaluated as shown in equation 2.38.[113]

$$\begin{aligned}
&\int {}^{3}\Psi_{SD}\frac{1}{r_{12}}{}^{3}\Psi_{SD}dr_1 d\omega_1 dr_2 d\omega_2 \\
&= \frac{1}{2}\left(J_{ab} - 2\int \psi_a(1)\psi_b(1)\frac{1}{r_{12}}\psi_a(2)\psi_b(2)dr_1\,dr_2 + J_{ab}\right) \\
&= J_{ab} - K_{ab}
\end{aligned} \tag{2.38}$$

The integrals J and K allow each electron to experience the average electrostatic repulsion of a charge cloud due to all the other electrons. The assumption that electron-electron repulsion occurs between an electron and a charge cloud rather than between all possible pairs of electrons is a major drawback of the Hartree-Fock method, which is discussed below. Equation 2.38 shows that for this wave function the classical Coulomb repulsion between the electron clouds in orbitals a and b is reduced by K_{ab} (electron exchange integral). J is the coulomb integral which represents the electrostatic repulsion between the charge clouds of orbitals a and b.[114] The probability of finding two electrons with the same spin in the same vicinity is reduced and a so-called 'Fermi-hole' is said to surround each electron. This property is unique to electrons of the same spin. If we consider the Slater determinantal wave function formed from different spins with the same assessment of interelectronic repulsion. The result of this shows that the exchange correlation is derived from the orthogonality of opposite spin functions.[115]

2.3.8 SELF-CONSISTENT FIELD

The electron density is calculated by squaring the wave function $\rho_j = |\psi_j|^2$. Since this equation cannot be used in the one-electron Hamiltonian, it lead to the introduction of the self-consistent field (SCF) proposed by Hartree in 1928.[116] The self-consistent field process guesses the wave function ψ for all the occupied molecular orbitals; also the variables being optimized are the coefficients. The solution to each differential equation 2.29 provides a new set of coefficients. The one-electron molecular orbitals[117] are then calculated with the new more accurate coefficients to determine the density matrix, which is comprised of coefficients. This process occurs iteratively until the difference between the newly determined one electron molecular orbital set and the immediately preceding set falls below some criterion.[118] This final set of one electron molecular orbitals is referred to as converged SCF orbitals. The sum of the individual operators *h* defines a separable Hamiltonian operator, which corresponds to the *non-interacting* system of electrons, since each electron sees a constant potential with which it interacts. The non-interacting Hamiltonian is not a good estimate of the true Hamiltonian because each *h* includes the repulsion of its associated electron with all of the other electrons i.e., h_i includes the repulsion between electron *i* and *j* but so too does h_j. If we summed up all of the one-electron eigenvalues for the operators h_i, we would count the electron-electron repulsion twice. We correct for this by using equation 2.39.[107]

$$E = \sum_i \varepsilon_i - \frac{1}{2}\sum_{i \neq j} \int\int \frac{|\psi_i|^2 |\psi_j|^2}{r_{ij}} dr_i dr_j \tag{2.39}$$

where *i* and *j* consider all the electrons, ε_i is the energy of the molecular orbital *i*, from the solution of the one-electron Hamiltonian defined by equation 2.31.

2.3.9 BASIS SETS

The basis sets used for Hartree-Fock (HF) are the mathematical function used to construct the HF wave function. From the theory above, we know that each molecular orbital in HF theory is expressed as a linear combination of basis functions, the coefficients for which are determined from the iterative solution of the SCF equations. The complete HF wave function is shown as a single Slater determinant

formed from the individual molecular orbitals. One cannot use an infinite basis set, thus much research has gone into identifying mathematical functions that allow wave functions to approach the HF limit as efficiently as possible. Basis functions should be in a form that is useful in some chemical sense. These functions should have large amplitude in space where the electron probability density is also large and *vice versa*.[107] Slater type orbitals (STOs) have spatial and electron spin orbitals, which suffer from certain limitations such as the computation time required to solve the two-electron integral, however high quality STOs have been developed for atomic and diatomic calculations and are mainly used in semi-empirical calculations. Modern molecular *ab initio* calculations utilize Gaussian orbitals, which differ from STOs[119] in that the exponent involves the square of the distance of the electron from the point on which the function is centered (usually an atomic nucleus). The general functional form of a normalized Gaussian-type orbital (GTOs) in an atom centered Cartesian coordinate is

$$\phi(x,y,z;\alpha,i,j,k)=\left(\frac{2\alpha}{\pi}\right)^{3/4}\left[\frac{(8\alpha)^{i+j+k}\,i!\,j!\,k!}{(2i)!(2j)!(2k)!}\right]^{1/2} x^i y^j z^k e^{-\alpha\left(x^2+y^2+z^2\right)} \tag{2.40}$$

where α is an exponent which determines the width of the GTO, and i, j and k are non-negative integers that controls the nature of the orbital on the Cartesian plane. When all three indices are zero, the GTO has symmetry and is named s-type GTO. When precisely one of the indices (p_x, p_y, p_z) is one, the function has axial symmetry and is called a p-type GTO. When the sum of the indices is two, the orbital is classified as a d-type GTO.[107] Gaussian functions are more computationally efficient, a single Gaussian function is a poor approximation of the wave function. To improve the approximation, one can use several Gaussians. Combining several Gaussians can give a better approximation to the wave function. The different combinations of GTOs and STOs are described in Chapter 3.

2.3.10 POLARIZATION FUNCTIONS

Polarization functions are generally added to heavy atoms but can also added to light atoms such as hydrogen and lithium. The only basis function located on a hydrogen atom is the minimal basis set, which is the minimum number of basis functions required to represent all of its electrons, and it would approximate to a 1s type atomic orbital.[120] The most common minimal basis set is STO-nG, it provides a rough estimate of the electronic structure. Adding a polarization function adds a p-function to the basis set and a p orbital on the hydrogen atom. As atoms get closer together, an atom's orbital might shift to one side or the other (polarization), if the shifting of orbitals is of particular importance in a study, it would be useful to add polarization functions, to better represent the actual chemical system.

2.3.11 DIFFUSE FUNCTIONS

Electrons and core electrons that are involved in bonding are confined within the framework of the molecule. Lone-pairs and unpaired electrons in a valence shell are not held as tightly and are further away from the nucleus than the core electrons. These expanded electron clouds occur in molecules that have highly electronegative atoms, such as oxygen, in anions and in excited states.[121] In order to represent this behavior accurately, diffuse functions are used since sometimes these electron rich areas of the molecule can change position. Diffuse functions are Gaussian functions with smaller values of α, which result in $e^{-\alpha r^2}$ to fall off more slowly as the distance (r) from the nucleus increases. A diffuse function is added with a single Gaussian for every valence orbital (the electrons that take part in chemical reactions are held in this orbital) on atoms such as oxygen and phosphorous.[120]

2.4 DFT

The Hartree-Fock theory discussed earlier illustrated a wave function by allowing it to be expressed as a Slater Determinant of one-electron orbitals; we can however work with a wave function that has a physical observable property in determining the

energy of a molecule. Going back to the Hamiltonian, we note that it depends on the position and atomic numbers of the nuclei as well as the total number of electrons. Since the Hamiltonian is used to describe the electrons, it would be useful to use the electron density as a physical observable, when integrated over all space; we obtain the total number of electrons.[122] DFT is therefore not based on a wave function but on an electron density function, which is indicated by $\rho\ (x, y, z)$ where ρ is the electron density over all of Cartesian space. The electron density is a measurable quantity. The DFT method is not considered a semi-empirical method due to the limited use of semi-empirical parameters and the possibility of finding its exact functional, makes the method *ab initio*.

The square of one electron wave function at any point x, is the probability of finding an electron in the volume *dx, dy, dz*. From a known density one could form the Hamiltonian, solve the Schrödinger equation, and calculate the wave function and energy eigenvalues. Since we are trying to find a relation between the ground state electron density of a molecule and its energy, some of the terms in the external potential can be considered in a classical manner where the electron density is treated as a charge distribution. For example the potential energy calculated between the density and the nuclei can be calculated as shown in equation 2.41.

$$V_{ne}\left[\rho(r)\right] = \sum_{k}^{nuclei} \int \frac{z_k}{\left(r - r_k\right)} \rho(r)\, dr \tag{2.41}$$

From the Hohenberg-Kohn theory we know that electrons interact with each other and with an external potential.[123] The external potential in a molecule is the attraction of the electrons towards the nuclei. The Hohenberg-Kohn Variational theory shows that this theorem is in correlation with MO theory in which the electron density obeys a variational principle and when the density is integrated one can obtain the number of electrons.[120] The variational principle states that the expected energy value must be larger and equal to the true ground state energy. We can therefore chose different densities and the ones that result in lower energies can be used. But we need to know when to stop choosing densities and whether the energy values associated to each density are closer to the correct energy.[120] Using the electron density we can determine the external potential, which is then used to obtain the Hamiltonian and hence the wave function and then solve the Schrödinger equation. But this approach does not provide any simplification over MO theory, since the final step is still the

solution to the Schrödinger equation. The most difficult aspect in the correct Hamiltonian is the electron-electron interaction. By considering the system as consisting out of non-interacting electrons (a fictitious electron gas), a molecule can be treated as a non-interacting electron density, associated with it an exchange and correlation term that can be related to this electron density by some functional. This ideas was first introduced by Kohn and Sham.[124]

2.4.1 SCF BY KOHN-SHAM

Kohn and Sham used a Hamiltonian that could be expressed as a sum of one-electron operators. The eigenfunctions are the Slater determinants of the individual one-electron eigenfunctions and eigenvalues are the sum of the one-electron eigenvalues. The technique here was to take as a starting point an *imaginary* system of *non-interacting* electrons that have for their overall ground state density the *same* density as some *real* system where electrons do interact.[107,120] The density is used to determine the position and atomic numbers of the nuclei; these quantities are the similar in the non-interacting and real systems. The energy functional can be expressed as[124]

$$E\left[\rho(r)\right] = T_{ni}\left[\rho(r)\right] + V_{ne}\left[\rho(r)\right] + V_{ee}\left[\rho(r)\right] + \Delta T\left[\rho(r)\right] + \Delta V_{ee}\left[\rho(r)\right] \tag{2.42}$$

where the terms represent the kinetic energy of the non-interacting electrons, the nuclear electron interaction, electron-electron repulsion treated classically, the correction to the kinetic energy due to the electron which interact with each other and all non-classical corrections to the electron-electron repulsion energy. For a system of non-interacting electrons, the kinetic energy is the sum of the individual electronic kinetic energies. An orbital expression for the energy in terms of the electron density may be rewritten as

$$E\left[\rho(r)\right] = \sum_{i}^{N}\left(\left\langle \chi_i \left| -\frac{1}{2}\nabla_i^2 \right| \chi_i \right\rangle - \left\langle \chi_i \left| \sum_{k}^{nuclei} \frac{Z_k}{|r_i - r_k|} \right| \chi_i \right\rangle\right) + \sum_{i}^{N}\left\langle \chi_i \left| \frac{1}{2}\int \frac{\rho(r')}{|r_i - r'|} dr' \right| \chi_i \right\rangle + E_{xc}\left[\rho(r)\right]$$

(2.43)

where N is the total number of electrons, the terms ΔT and ΔV_{ee} are taken together as a term E_{xc} which is referred to as the exchange correlation functional. Incorporated in this term are the effects of quantum mechanical exchange and correlation, as well as

the correction for the classical self-interaction energy and for the difference in kinetic energy between fictitious non-interacting system and the real one. Determining the Kohn-Sham orbitals, we express the orbitals within a basis set of functions and solve the individual orbital coefficients by solution of a secular equation similar to its use in HF theory discussed earlier, with the $F_{\mu\nu}$ replaced by $K_{\mu\nu}$ and is defined as

$$K_{\mu\nu} = \left\langle \phi_\mu \left| -\frac{1}{2}\nabla^2 - \sum_k^{nuclei} \frac{Z_k}{|r - r_k|} + \int \frac{\rho(r')}{|r - r'|} dr' + V_{xc} \right| \phi_\nu \right\rangle \tag{2.44}$$

Solution of the Kohn-Sham process is an iterative procedure since the calculation of the updated density requires the density itself. While Hartree-Fock is an approximate theory, density functional theory is exact since one is able to solve the relevant Kohn-Sham equations exactly. Hartree-Fock theory or MO theory optimizes a wave function whilst DFT theory optimizes electron density. The B3LYP model used in this study is a hybrid functional, which was introduced by Alex Becke.[125] This gradient-corrected method incorporates a Hartree-Fock exact exchange functional allowing for the improvement of molecular properties such as bond lengths and atomization energies.[126] It is termed a hybrid functional since it incorporates HF and DFT theory.

2.5 SCC-DFTB

SCC-DFTB is an approximate quantum semi-empirical method derived from density functional theory. DFTB is a method that expresses the electron density as a reference density with a small fluctuation. The density is then held fixed while the energy is minimized only in relation to the shape of the Kohn-Sham orbitals. To represent the density fluctuations in a simple way as the tight-binding method it is written as a superposition of atom-like contributions, which decays along the distance from the atomic center. By choosing a density we can optimize the orbitals without the need of computing the electron density again. This approach of choosing an electron density is *non-self-consistent*. The secular equation is solved to obtain the KS orbitals basis set coefficients, which are used, in the variational minimization or SCF cycle. The accuracy of the DFTB method becomes poorer when a more uneven charge balance controls the bonds between atoms such as heteronuclear molecules. In

order to correct for this, a second order contribution was introduced in SCC-DFTB. This method is an extension of the DFTB method since it helps describe systems with electronic character better.[127] The electron density is written as a superposition of neutral atom electron densities and fluctuations. The fluctuations allow for charge to be transferred between atoms whereas in the DFTB process this is not possible.

SCC-DFTB was the semi-empirical method applied in this study with hydrogen bond and dispersion corrections. A semi-empirical technique was used since MM cannot describe quantum effects such as changes in electronic structure or charge transfer as discussed earlier.[128] MM does not consider polarization affects either. Semi-empirical techniques are used to overcome some of these limitations in MM such as the application of H-bond and dispersion correction.[129] These corrections are added because the structure and function of biomolecules are dominantly affected by non-bonded interactions like dispersion and hydrogen bonding.[129] SCC-DFTB also predicts barrier heights for conformational change more accurately when coupled with the FEARCF method compared to other semi-empirical techniques.[3]

From the discussion above we can establish that DFT optimizes electron density whilst HF theory optimizes a wave function. If we want to determine a particular property of a molecule and would like to know which is the best method to use, we would need to know how that property depends on the electron density or the correct quantum operator to determine that property using a wave function. For example consider the total energy of interelectronic repulsion; if we chose to use DFT and we had the exact density for the system, we however do not know the exact exchange correlation (E_{xc}) functional and therefore we cannot compute the exact interelectronic repulsion. If we chose to use HF theory (not the single determinant HF, since electron-electron correlation is not completely accounted for) we would have an approximate to the wave function in which we could evaluate the expectation value for the interelectronic repulsion operator to determine the energy. Both these methods can be computationally expensive and time increases with size of the system, for the DFT theory it is N^3 where N is the number of basis functions used to represent the KS

orbitals.[107] Choosing a method would depend on the system and the level of accuracy one would like to obtain.

In this study, a DFT method was chosen for a system size of 27 atoms, with relatively reasonable computational time. Diffuse functions were added to the basis set to account for any positional change in electron density clouds as the conformation of the monosaccharide ring changed. The exact basis functions chosen and their applicability to carbohydrates are discussed in the following chapter.

CHAPTER 3

COMPUTATIONAL AND ANALYTICAL METHODS APPLIED TO CARBOHYDRATES

Computational characterization of carbohydrates and their conformational analysis, has posed a challenge over the years.[130] The conformational analysis of carbohydrates can be quite complex since they exhibit numerous conformations that coexist in solution at room temperature. The numerous conformations arise from hydrogen bonding, primary and secondary alcohol rotation, ring puckering, glycosidic linkage torsion angle rotation[131] and ring opening[132]. Carbohydrates exist as a conformational ensemble in solution which places setbacks on techniques such as nuclear Overhauser intensities (NOE) that use solution derived data to deduce the conformation.[99] This technique provides the time averaged properties of a molecule, which is suitable for relatively rigid systems, since the NOE intensities can be related to inter-proton distances which are used to derive reasonable conformations. However, carbohydrates are extremely flexible, use of this method may lead to the inaccurate analysis, since virtual conformations may be generated which may not physically exist.[133]

3.1 CARBOHYDRATE FORCE FIELDS

The methods used to study carbohydrates and their conformations must consider both spatial and temporal properties. The success of these methods depends greatly on their ability to represent the inter-atomic properties of the molecule as closely as possible to the actual system.[99] A force field as discussed in Chapter 2 is required to describe the inter-atomic properties of a molecule. There are different force fields that have been applied to carbohydrates, each with their advantages and drawbacks. The force field known as the hard sphere exo-anomeric effect (HSEA)[134,135], assumes that the conformation of an oligosaccharide is governed by van der Waals interactions and an added torsion potential, the added torsion potential is meant to account for the exo-anomeric effect. This force field ignores electrostatic effects such as hydrogen bonding and dipolar interactions. Another set of force fields accounts for the

energetic contributions from bond stretching, angle bending, torsional rotation and non-bonded interactions such as the common macro-molecular force fields (see Section 2.2.2). The MM2 or MM3 force field have been used to predict the conformations of mono- and disaccharides, which are more sophisticated than the common macro-molecular force fields.[136,137] Since refined mathematical expressions are used to describe bond stretching and angle bending more efficiently, it has accurately reproduced details of molecular structure, including ring deformation and bond distance changes arising from the anomeric effect.

3.1.1 AMBER AND GROMOS FORCE FIELDS

The Amber force field has been edited to improve the reproducibility of the molecular properties of α-linked carbohydrates.[138] The AMB99C force field considers the primary alcohol rotation, glycosidic linkage rotation and the anomeric effect. This force field was designed for the study of α-1-4 linkages in saccharides.[139] The GROMOS carbohydrate force fields based on the original parameter set (43A1 or 45A3) have shown an incorrect description of the anomeric effect, ring conformations and incorrect dihedral angles distributions in saccharides.[140-142] The revised GROMOS force field 45A4 included these aspects in the parameterization and was only validated for a small set of mono- and disaccharides in solution.[142] The GROMOS 53A6$_{GLYC}$ force field is an improved parameter set for hexopyranoses and includes refinements of dihedral angles that determine pyranose ring conformations.[143]

3.1.2 CARBOHYDRATE SOLUTION FORCE FIELD

A computational model must adequately predict the thermodynamic and transport properties of saccharides in order to interpret their conformational function in solution accurately.[45] The CSFF method developed specifically for carbohydrates and their primary alcohol rotation parameterization by Kuttel et al.[45] addresses some of the issues related carbohydrate conformational sampling in solution. Saccharides have conformational freedom about the glycosidic linkage, primary and secondary alcohols. The primary alcohol group has three staggered conformers (*tg*, *gt* and *gg*,

discussed in Chapter 1). The factors affecting the conformational preferences of the primary alcohol group are stereoelectronic effects, steric 1-4 interactions, hydrogen bonding and solvent effects.[144] An accurate carbohydrate force field for molecular dynamics simulations should reproduce the experimental ratios of each primary alcohol rotamer and the experimental time scale between rotations. Therefore, the frequency of transitions between the primary alcohol rotations, which is controlled by the free energy barrier heights between conformational minima, is addressed in this method. The CSFF method modified the parameter sets of the PHLB force field, which exhibit infrequent conformational transitions of the primary alcohol group, to alter the relative free energies of the staggered conformers. The method produces primary alcohol equilibrium distributions and rotational frequencies that are in agreement with experimental observations for saccharide units.[45]

3.1.3 CHARMM FORCE FIELDS

CHARMM includes a set of all-atom force fields for proteins, nucleic acids, lipids and carbohydrates that are compatible with each other. With the advent of longer residue name codes in CHARMM and a great resurgence of interest in glycobiology more specifically a significant extension upon the CHARMM carbohydrate force field has been made by MacKerell.[145] More specific force fields have been parameterized by MacKerell for glycosidic linkages in carbohydrates involving furanoses[146] and pyranoses[146,147] as well as their monosaccharide derivatives[97,147]. The monosaccharide units that are important components of eukaryotic glycans such as glucuronic acid, iduronic acid, xylose, fucose, *N*-acetylglucosamine have been specifically parameterized for their function in carbohydrate-protein modeling.[148] Parameter sets were also developed for aldose, ketose linear carbohydrates and sugar alcohols.[149] The parameter sets for the pyranose monosaccharide unit's considers the rotation of the primary alcohol group, exocyclic hydroxyls and conformational change with the parameters validated against experimental data (thermodynamic quantities heat of vaporization, molecular volume, free energy of aqueous solvation and infrared vibrational frequencies).[97]

3.2 CONFORMATIONAL ANALYSIS

Finding the molecular structure with the lowest energy can be a tedious process when studying carbohydrates, since energy surfaces for molecules that have many internal degrees of freedom, have more than one local minimum. The algorithms that minimize the energy in MM or QM will proceed with the starting conformation and following the energy gradient downhill to optimize into the closest local minimum on the energy surface. Monosaccharide units have various structural orientations that can affect the local minima. The cyclic portion of the molecule can librate, twist and flip and thus change conformation for example from one chair conformation to another - $^4C_1$ to $^1C_4$.[150] The three staggered conformers of the primary alcohol group are likely to correspond to a local minimum. Each exocyclic or secondary hydroxyl could also orient in one of three staggered conformations. A systematic rotation of the primary alcohol with a minimization of the structure at each point would have to consider alternate hydroxyl hydrogen orientations (each possible staggered rotation of the secondary hydroxyls with each rotation of the primary alcohol) that may result in lower energy. In disaccharides, the rotation of the glycosidic linkage torsion angles must be considered and is also likely to correspond to a local minimum.[131]

The rotation of the hydroxyl groups and the glycosidic linkage with its corresponding energy were studied for β-maltose by Momany.[151] The conformational behavior of the glycosidic bond is governed by the interactions of the hydroxyl and functional groups from each ring as well as interaction between rings across the glycosidic link. Obtaining the global minimum energy for maltose required an exhaustive conformational search of the hydroxyls and dihedral angles used to define the orientation of the glycosidic bond.[151]

3.3 POTENTIAL ENERGY SURFACES AND VOLUMES

The energy associated with each structural variation in the monosaccharide unit can be plotted and interpreted as a potential energy surface, with areas of maxima and minima and the related structures. A potential energy surface (PES) with more than two variables can be plotted as a potential energy volume (PEV), with three dimensions. The optimizations of small molecular systems in vacuum can assist in

understanding the effects of geometry and conformation on enthalpically based macroscopic properties, since it has been reasoned that the PEV and stationary points on the PEV can be linked to the macroscopic behavior of molecular systems.[152] This approach in conformational analysis often does not have the essential information when investigating complex mechanisms[153] such as the hydrolysis of cellulose in the enzyme binding pocket, which involves hydroxyl rotation and ring deformation. This is due to chemical systems and molecules being dynamic, in order to obtain information on complex transition states and products, the direct sampling of phase space is required. The structures obtained from each optimization on the PEV are not comparable to the number of configurations available to a system when the thermal energy is increased, at a specified set of structural variations (primary alcohol rotation, ring pucker etc.) or reaction coordinate.[153]

Investigating a system using multidimensional PES and considering every degree of conformational freedom will require lengthy amounts of computational resources and time and in addition is not useful in describing the dynamical behavior and properties of the molecular system. A molecular dynamics simulation includes fluctuations in conformation due to thermal energy and can be used to generate molecular ensembles from which it is possible to derive properties of complex systems that are difficult to calculate from the PES. Casting the ensemble statistics in terms of coordinates interest, it is possible to calculate information on complex transition states and intermediates involving enzymes[154], protein folding[155-157], bond breaking and bond forming from a quantity termed the free energy. The free energies of a system are investigated as a function of one or two structural variations or reaction coordinates using non-Boltzmann dynamic calculations, which are used to generate a potential of mean force (described later in the chapter). The interpretation of the system is considered in terms of these reduced reaction coordinates and the information from the Boltzmann sampling (statistics) of the other conformational variable such as exocyclic hydroxyl rotation and primary alcohol rotation are implicitly included.

3.4 FREE ENERGY

Free energy is obtained by subtracting the product of the absolute temperature and entropy from the internal energy of a system. Free energy is considered as the most central quantity in thermodynamics, as it represents the energy of a system that is available to be converted to carry out work. The free energy is expressed as the Helmholtz function, *A*, or the Gibbs function, *G*.[83] The Helmholtz free energy is the energy available to do work for a system that has a constant number of particles, temperature and volume (NVT ensemble) whilst the Gibbs free energy is the free energy for a system that has a constant number of particles, temperature and pressure (NPT ensemble).[158] A Gibbs free energy change can be thought of as an enthalpy change adjusted by a temperature-weighted entropy change as shown in equation 3.1.

$$\Delta G = \Delta H - T\Delta S \tag{3.1}$$

The $T\Delta S$ term is a minor contributor to the Gibbs free energy at room temperature, but becomes more significant at sufficiently high temperatures.[120]

The Helmholtz free energy at constant NVT is calculated by the internal energy (*U*) which is adjusted by the temperature (*T*) and entropy (*S*) of the system, see equation 3.2.[120]

$$A = U - TS \tag{3.2}$$

Free energy can not be accurately determined from a standard sampled molecular dynamics simulation, since these simulations do not sufficiently sample those regions of phase space that make significant contributions to free energy. Molecular dynamics sampling seeks out areas of phase space that are low in energy, therefore these simulations will not sample regions of phase space that are high in energy.[83] It is however, possible to calculate the differences in quantities such as the Gibbs free energy and Helmholtz free energy.

3.4.1 CALCULATING FREE ENERGY VARIANCES

We will now consider calculating the Helmholtz free energy difference of two states A and B using thermodynamic perturbation and thermodynamic integration.[83]

3.4.1.1 THERMODYNAMIC PERTURBATION

The difference in entropy properties between the two systems A and B can be determined from an ensemble average.[82]

$$\left\langle \Delta A_{A\to B} \right\rangle_M = -kT \ln \left\langle e^{-(E_B - E_A)/kT} \right\rangle_M \tag{3.3}$$

The transformation from states A to B is sectioned into several intermediate steps which is described by a parameter called λ and the total energy is given as the sum of changes in each step. The perturbation from A to B is generally tested by perturbing in the reverse direction as well. Obtaining a reliable estimate of the free energy requires several independent simulations with the analysis of the perturbation steps at each simulation.[82] Calculating the free energy differences by means of equation 3.3 is called thermodynamic perturbation[159,160].

3.4.1.2 THERMODYNAMIC INTEGRATION

The total energy which is described as a function of λ results in the partition function and free energy given as a function of λ, see equation 3.4.[82]

$$A(\lambda) = -kT \ln Q(\lambda) \tag{3.4}$$

Differentiating the equation 3.4 with respect to λ, yields

$$\frac{\partial A}{\partial \lambda} = -\frac{kT}{Q}\frac{\partial Q}{\partial \lambda} = \frac{\partial E}{\partial \lambda} \tag{3.5}$$

Substituting the right hand side by an ensemble averages and integrating over λ results in equation 3.6.[82]

$$A(1) - A(0) = \int_0^1 \left\langle \frac{\partial E(\lambda)}{\partial \lambda} \right\rangle_M d\lambda \tag{3.6}$$

The left hand side of equation 3.6 is the free energy difference, this approach of calculating free energy is termed thermodynamic integration[83,161].

Molecular dynamics simulation are unsuitable in accessing all of conformational space, since some systems have multiple degrees of freedom, with many energy barrier heights. The conformers only accessible at high energies are sampled infrequently or not at all, leading to the sampling of a small set of minima. However, the molecular system can be persuaded to visit areas of conformational space that were not previously sampled and not be confined to small number of low energy regions.[162] The next set of topics look at the use of a biasing potential, to persuade the sampling of new areas of conformational space.

3.4.2 POTENTIAL OF MEAN FORCE (PMF)

We will now consider how the free energy changes with respect to some inter- or intramolecular coordinate such as the torsion angle of a bond within a molecule, the distance between two atoms or a dihedral torsional angle of three bonds within a molecule. The free energy change calculated based on the probability distribution along a particular coordinate is referred to as a potential of mean force.[163] Unlike the modifications that occur in the thermodynamic perturbation method, the potential of mean force[164] is calculated for a physical reason, such as a reaction coordinate (a subspace of conformationally relevant degrees of freedom). A simple example of a PMF would be the free energy change as the distance (r) between two atoms. We could calculate the potential of mean force from the radial distribution function with the following expression for the Helmholtz free energy[83]

$$A(r) = -k_B T \ln g(r) + cons \tag{3.7}$$

The constant is chosen such that the most probable distribution correspond to a free energy of zero, the distances that are sampled the most will have low free energies. The potential of mean force can vary with several degrees of magnitude of $k_B T$ with a change in r leading to inaccurate estimates of the PMF. Using a technique such as umbrella sampling can rectify this problem.

3.4.3 UMBRELLA SAMPLING

This type of sampling serves to overcome the problem mentioned above by modifying the potential function such that unfavorable areas of phase space are adequately sampled. Umbrella sampling can be used with MD simulations.[83] The modification of the potential function is written as follows

$$V'\left(r^N\right) = V\left(r^N\right) + W\left(r^N\right) \tag{3.8}$$

where $W(r^N)$ is a weighting function[165], for configurations in phase space that are away from the equilibrium state designated by r_0^N, the weighting function will be a significant amount, therefore a simulation that utilizes the potential function will be biased away from configurations in the equilibrium state. Let us consider the PMF for rotation using umbrella sampling, of the central C-C bond of ethane. The energy barrier between the rotamers eclipsed and staggered is significantly high, but by applying a modified potential to the simulation, there was a shift in expected populations of each rotamer, making the high energy rotamer more accessible.[166] This method requires that the biasing potential is initially guessed and then refined iteratively.

3.4.4 LOCAL ELEVATION UMBRELLA SAMPLING (LEUS)

The local elevation method utilizes a memory-dependent potential energy term in the MD simulation to prevent the resampling of areas in phase space that were already sampled, leading to the sampling of new configurations available in phase space.[162] The LEUS method combines the local elevation conformational searching and the umbrella sampling conformational sampling techniques for molecular dynamics. In this method, the optimized biasing potential is constructed via an initial local elevation searching phase, for a meaningful set of reaction coordinates, followed by umbrella sampling.[167] The LEUS technique has been applied to the study of relative free energies and interconversion barriers of glucose ring pucker in solution.[168] This technique does not iteratively preoptimize the biasing potential like the Adaptive Biasing Force method and is used because it is an efficient, versatile technique for improving the sampling in MD simulations.[168]

3.4.5 ADAPTIVE BIASING FORCE (ABF)

Unlike the umbrella sampling method the biasing force in the ABF[169-171] method is estimated from the previously sampled configurations of the system and is continuously updated as the simulation continues.[171] Since this biasing force is continuously updated, it can be referred to as adaptive because it changes as the simulation progresses and utilizes all the statistics obtained thus far in the simulation to improve the sampling. The biasing force in the ABF[172] method is estimated as soon as enough sampling points are obtained in a given bin. A bin would refer to a part in phase space, with phase space being all the possible configurations for a coordinate under investigation.[83]

As discussed in Section 3.4 in normal MD simulations, significant parts of phase space are not easily accessible.[92] These areas of phase space are not sampled since they are much greater in free energy than the energy of minima. Accessing areas of phase space of high energy in MD simulations such as transition states are only possible by biasing the sampling away from previously sampled areas by using a biasing potential or force, the FEARCF method is one such method that samples conformers higher in free energy by the use of a biasing potential .[92]

3.5 FREE ENERGY FROM ADAPTIVE REACTION COORDINATE (FEARCF)

The initial application of the FEARCF method occurred in 1999, where the two-dimensional conformational potential of mean force for a disaccharide unit was determined in solution.[44,92] The method was then combined with quantum classical dynamics (QM/MM), which was used to simulate chemical reactions free energy surfaces.[173] The flat histogram approach was implemented in the FEARCF method to achieve equal visits to each state defined in the reaction coordinate. The flat histogram technique previously used weighting factors which were added to the state probabilities,[174,175] while the more recent use of this method involved the determination of the partition function at multiple steps during the simulation.[176,177] The FEARCF method was the earliest technique to implement the flat histogram

method in MD simulations and it sought to develop the partition function using an ensemble of simulations.[92,178]

The FEARCF method is a non-equilibrium biased dynamics.[92,153] In the FEARCF method one tries to obtain equal sampling across reaction coordinate phase space in order to obtain the best guess of the pmf. The potential of mean force, a biasing potential is related to the probability density (sampling of phase space) by equation 3.9, where k_b is the Boltzmann constant[92] with temperature in Kelvin (K) and ξ is the reaction coordinate for ring pucker. The potential of mean force (W) and probability density (P) are unknown at the beginning of the simulation.[178] The pmf is derived in a canonical ensemble with constant number of atoms, volume and temperature of 298K,

$$W(\xi) = -k_b T \ln P(\xi) \tag{3.9}$$

Initially the potential of mean force is zero. From the initial simulation the probability distribution is calculated by looking at the sampling, this resulting probability distribution is used as a first guess for W from which an improved biasing function $U(\xi)$ for the next simulation is obtained. This process continues iteratively until the entire pucker conformational space is adequately sampled. In order to overlap the histogram data obtained from the sampling of phase for multiple simulations the weighted histogram analysis method is used. To traverse barrier heights greater than $3k_b$T, one applies a biasing function. The best guess for the biasing function is the inverse of the pmf $W(\xi)$ shown in equation 3.10.[178]

$$U(\xi) = -W(\xi) \tag{3.10}$$

The potential of mean force for ring pucker, $W(\xi)$ is calculated as a function of the three dimensional coordinate set ξ.

$$\xi = \theta_0 + \theta_1 + \theta_2 \tag{3.11}$$

The three dimensional coordinate set consists of the three pucker angles obtained from the triangular decomposition method see equation 3.11. At each step in the simulation the biasing force (F_i) for θ_i are applied to atoms involved in the rotatable plane is calculated from the gradient of the biasing function for three independent

reaction coordinate shown in equation 3.12, where F is calculated as a partial derivative. A multidimensional cubic-spline interpolation is used to calculate the biasing force, $F(\xi)$ for the reaction coordinate. The biasing force is applied to the reaction coordinate to bias the next simulation's trajectory away from previously samples parts of phase space. This force also assists the system in sampling areas of phase space that have barrier heights that are larger than those thermally accessible in a MD simulation.

$$-\frac{\partial U(\xi)}{\partial \theta_i} = F_i = \frac{\partial W(\xi)}{\partial \theta_i} \tag{3.12}$$

The probability density is a three dimensional grid which keeps a running tally of the sampling of pucker conformational phase space, derived from the history of simulations until that point.[178] From the free energy surface, free energy pathways can be obtained if the reaction coordinate is sampled uniformly or has converged. A converged free energy surface occurs when all of phase space is adequately sampled.

3.6 REACTION COORDINATE

In order to apply the FEARCF method a reaction coordinate for ring pucker has to be defined and calculable. A simple example for reaction coordinate mentioned above is the bond distance between two H atoms. At each dynamic step in a molecular dynamics calculation for the FEARCF method, the reaction coordinates are evolved i.e. the reaction coordinates are calculated and recorded at each step during dynamics.

3.6.1 RING PUCKER

As introduced in Chapter 1, saccharide units in addition to the conformational freedom about their primary alcohols, have a special degree of conformational freedom where the ring deforms and is referred to as ring pucker. The ring can flip, twist and bend which results in 38 different canonical states for a six-membered ring such as β-D-methyl glucose shown in Figure 3.1. Amongst these 38 canonical states[179] are chairs, boats, skew-boats, envelopes and half chairs with the unfavorable planar conformer rarely considered.

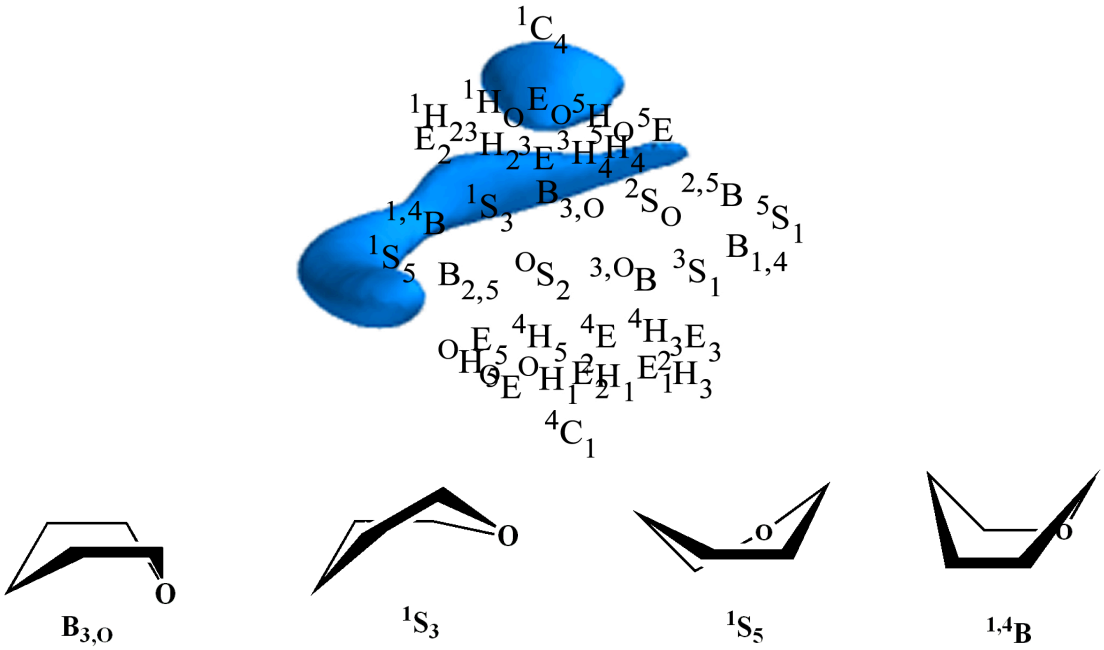


Figure 3.1 The three-dimensional Hill Reilly sphere for pyranose rings with three distinct hemispheres, from $^{4}C_{1}$, the lower or first hemisphere with conformers $^{4}H_{5}$, ^{4}E and $^{4}H_{3}$, the equator with boats and skew-boats $B_{3,O}$, $^{1}S_{3}$ and $^{1}S_{5}$ and the upper or second hemisphere with conformers $^{3}H_{4}$, $^{5}H_{4}$ and ^{5}E (adapted from[15])

Ring pucker depends on the position of the atoms defined in the ring, resulting in six degrees of freedom. Utilizing all six degrees of freedom will be very complex, hence the use of a reduced coordinate set initially introduced by Kilpatrick, Pitzer, and Spitzer[180] in their 1947 paper where they worked on out-of-plane deformations of cyclopentane.

3.6.2 THE CREMER - POPLE METHOD

The Cremer-Pople set is a popular tool for identifying the conformation of six and five membered rings and describes the deformation of a *N*-membered ring using normal out-of-plane modes of analysis.[181] Cremer and Pople introduced the analysis and characterization of pucker for six membered molecular rings. The method can be used without approximation to any cyclic molecule when only the nuclear positions of the atoms in the ring are given. This method however does not help to determine the change in conformation, which can occur as a result of external forces.[181] Although it has been utilized to calculate the conformational landscape of ring pucker for glucose using metadynamics[15], the validity of Cremer-Pople coordinates as a

collective variable has been questioned.[182] Following this method was the triangular tessellation method proposed by Hill and Reilly. It describes the pucker of a *N*-membered monocyclic ring using *N*-3 parameters.[183,184]

3.6.3 THE TRIANGULAR TESSELLATION METHOD

The triangular tessellation of molecular rings or triangular decomposition method as it has been recently renamed[183], is a reduction of the six degrees of freedom required to describe a six membered monocyclic ring. This representation is attractive since the mathematical difficulty of solving equations is reduced, whilst the parameters still retain all the information required to describe the ring conformation. The six-membered monosaccharide units looked at in this study may be described as moving through 38 different ring conformations.[179] Quantifying the change in conformation requires a geometric breakdown of the ring into a reference plane formed by the three atoms in a plane with *N*-3 flaps that have some angle of elevation to the plane. Each of the angles in the flap are calculated by first defining a reference plane.[183] This technique is suitable for a force based approach such as the FEARCF method because one can understand easily how a force applied to the reaction coordinate can lead to geometrical change of the ring.

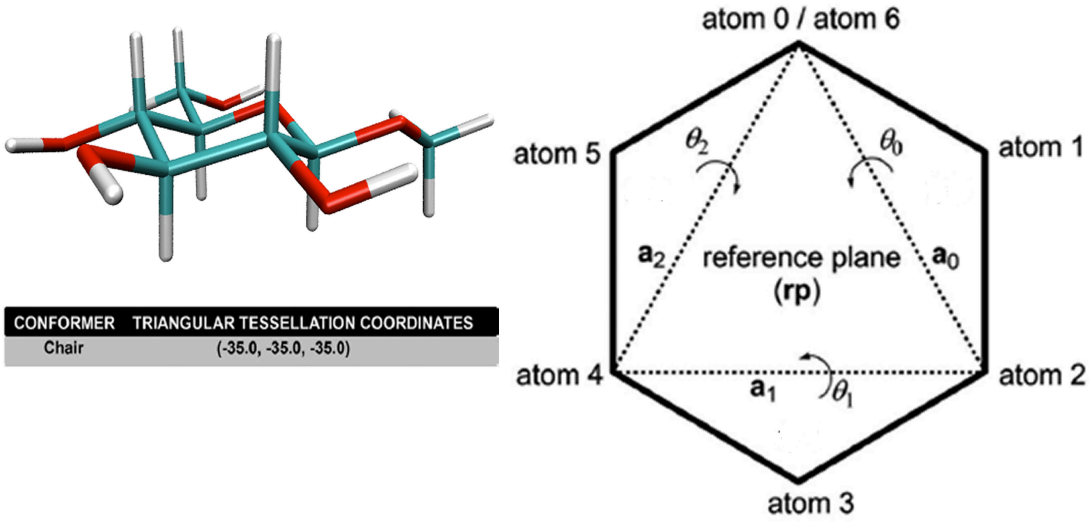


Figure 3.2 A tessellation of a pyranose ring into three flaps and a chair conformation with its associated triangular tessellation angles that each of the ring flaps make with the reference plane (inserted from[185])

Three atoms on the same plane as depicted in Figure 3.2 define the reference plane. The Cartesian coordinates of each x_i atom represented by atom 0-5 are used to calculate the axes of puckering a_i, with the numbering starting at zero as shown in equation 3.13.[183]

$$a_i = x_{2(i+1)} - x_{2i} \tag{3.13}$$

Using two of the axes of puckering, one can calculate the vector normal to the reference plane, by calculating the cross product between the axes.[183]

$$n = a_1 \otimes a_0 \tag{3.14}$$

Calculation of the bond vector that represents each of the bonds between atoms can be done using equation 3.15.[183]

$$r_i = x_{i+1} - x_i \tag{3.15}$$

The bond vectors can be used to calculate the atoms orientation vector, p_i relative to the plane by solving the cross product between two bond vectors. The vector q_i is orthogonal to p_{2i+1} and a_i, which is used to calculate the angle of pucker.[183] The angle of intersection between the vectors q_i and n is equal to $\pi/2 - \theta_i$, therefore θ_i as shown in equation 3.16 is equal to the cosine function of q_i and n.[183]

$$\theta_i = \frac{\pi}{2} - \cos^{-1}\left[\left(q_i \bullet n\right) \bullet \left(\left\|q_i\right\| \bullet \left\|n\right\|\right)^{-1}\right] \tag{3.16}$$

3.6.4 RING PUCKER IN FEARCF

The triangular tessellation method is used for β-D-methyl glucose with the reference plane defined by atoms C2, C4 and O5.[92] The theta angles θ_0, θ_1 and $\theta_2 \in [-90^0, 90^0]$ are the angles between atoms C4-C5-O5, O5-C1-C2, C2-C3-C4 and respectively; θ_0 describes the angular movement of the carbon (C5) to which the hydroxymethyl or primary alcohol group is attached, θ_1 describes the movement of the anomeric carbon to which the OR (R=monosaccharide unit or glycan) group is attached, θ_2 describes the movement of a carbon (C3) to which a secondary alcohol is attached.[3] From the sampling of each theta angle one can obtain information on minimum free energy

pathways, coordinate data for SCF analysis, lowest and highest free energy conformers.

3.7 BASIS FUNCTIONS

As discussed In Section 2.3.9 basis sets consist of a set of functions that are combined in linear combinations to create molecular orbitals. During most chemical reactions, it is the valence electrons that interact and take part in bonding, representing these electrons with more basis functions can increase the accuracy of the result. A mixture of basis functions allows for the electron density to adjust and delocalize electrons according to the molecular environment.[186] The Pople basis sets are examples of mixed basis functions and include 6-31G*, 6-311G, 6-311G++ etc.[187-189] The first digit indicates the number of primitive Gaussian orbitals that comprise each core atomic basis function. The asterisk indicates polarization functions, whilst the plus sign indicates that diffusion functions are added.

3.7.1 BASIS SETS APPLIED TO CARBOHYDRATES

Choice of basis sets requires knowledge on the chemical system, its conformational change, size and electronic structure. Disaccharide units contain more than 20 heavy atoms (carbon, oxygen, nitrogen) as well as hydrogen atoms with the ability to hydrogen bond.[190] This type of electrostatic interaction requires the electron correlation to be explicitly defined in the basis function.

The application of Hartree-Fock theory to saccharide units shows systematic errors caused by its deficiency in accounting for Coulomb-type electron correlations. Although, coupling HF theory with mixed basis sets such as 6-31G* and cc-pVDZ could give relative good energies for carbohydrate conformers.[191] Correlated methods like MP2, MP3 and MP4 treat coulomb electron correlation explicitly and are generally coupled with mixed basis sets, which give reliable equilibrium geometries and energies.[192] DFT methods treat electron correlation more efficiently than HF theory and correlated methods[193], with the most often used density functional for the study of carbohydrate conformations[194-198] has been B3LYP. The addition of diffuse

functions on the heavy and light atoms and polarization functions with a large basis set (6-311++G**) have been shown to be necessary when applying B3LYP in the study of carbohydrate conformers.[151,198-200] Although Csonka[195] suggested that 6-31+G* and 6-31+G** basis sets are sufficient for geometry optimization calculations for carbohydrates.

The basis sets that have been applied to disaccharide units are B3LYP/6-31G* and B3LYP/6-311++G**[201], with particular attention to conformational change around the glycosidic bond.[151] Since the rotation and stereochemistry of the glycosidic bond has an effect on the biological function of saccharides. Basis sets such as MP2/6-31G* and 6-311++G** have been employed to study the staggered orientations for rotation about the glycosidic bond, with the anomeric and exo-anomeric effects influencing the conformer that is preferred.[51] The basis sets B3LYP/6-311++G** and B3LYP/6-31G* have also been used to investigate transition states and intermediates on the reaction pathways of glycosyltransferases[202] and the catalytic mechanism of inverting *N*-acetylglucosaminyltransferase.[203] The B3LYP/6-31G* and B3LYP-6-31+G** basis sets have been used to study primary alcohol orientation in solvation[53] since these basis sets have been shown to produce accurate structures and energies for monosaccharide units.[204] The B3LYP/6-31+G* basis set was found to be more efficient than the B3LYP/6-31G*, with the C-O-H angle, hydrogen bonding distances and energies being better described for α- and β-glucopyranose.[200]

3.7.2 SELECTION OF BASIS SET

The MP2 level of theory was applied to the monosaccharide system with a range of basis functions (6-311++G, 6-311++G*, 6-31++G). The energies and electron density trends were compared to B3LYP/6-311++G(d,p). The B3LYP/6-311++G(d,p) basis function provided energies and electron density trends that were similar to the MP2 level of theory. The B3LYP basis set was therefore selected since the quality of the results was not comprised, with minimal computation time. Similar B3LYP basis sets have been used to probe carbohydrate transition state structures and the addition of diffuse functions have been shown to describe the C-O-H angle and energies

better.[200,202] Optimization of the geometry occurred with convergence of the SCF cycle, even at C1-O1 bond lengths ≥ 2.00 Å.

3.8 SIMULATION DETAILS FOR β-D-METHYL GLUCOSE

The β-D-methyl glucose molecule was built in CHARMM, using the CHARMM general all-atom carbohydrate force field.[96] For all simulations in CHARMM, the SCC-DFTB[3] semi-empirical potential was used. For the C1-OMe extended bond length simulations, a harmonic constraint was placed on the C1-OMe bond length and *in vacuo* dynamics were conducted in CHARMM 35[205] for 20 iterations with exception of the C1-OMe bond length > 1.90 Å. At each iteration eight FEARCF calculations with an initial biasing potential of zero with 0.5 ns in length were run using velocity-Verlet dynamics at 298.15 K with group cutoffs of 10, 12 and 14 Å and a random seed generator. This was done to further increase the reaction coordinate sampling; the batches of eight simulations were started from various pucker conformations (initially all $^4C_1$ and then the last conformer sampled from the previous run). Force switching was used to treat the electrostatics and the van der Waals potential were shifted to account for long distance non-bonded interaction discontinuities.[3] SCC-DFTB calculations were run with the mio-0-1 parameters and an improvement to hydrogen bonding was included with the HBON keyword. The sampled areas of phase space were collected to obtain the probability density $P(\xi)$, from which the first guess for the potential of mean force $W(\xi)$ were calculated and the biasing function $U(\xi)$ applied to the next batch of eight simulations. This iterative procedure was accomplished 20 times, totaling a sampling of 80 ns ($20 \times 0.5 \times 8$) for a C1-OMe bond length of 1.40 Å to 1.90 Å. For a C1-OMe bond length of 2.00 Å to 2.40 Å, 8 iterations were accomplished.

3.9 DFT ENERGIES OF SELECTED CONFORMERS

Canonical conformers ($^4C_1$ and the first latitude through pucker phase space, Figure 3.1) of β-D-methyl glucose with a C1-OMe bond length of 1.55 – 2.40 Å were extracted from FEARCF simulation trajectories. A single set of extracted coordinates, one closely representative of the actual canonical conformer was used in to obtain the global minimum by running minimization calculations for each rotation of the hydroxyl group. The minimization was done using conjugate gradient and SCC-DFTB. The minimum energy structure from these minimizations was then optimized using B3LYP/6-311++G(d,p). The lowest energy structures from these optimizations were then used in the bond scan. This approach allowed one to obtain the global minimum for each pucker conformation. Each conformer was then optimized with B3LYP/6-311++G(d,p) at each bond scan point.[206] The high-energy conformers were restrained across the three dihedrals in the ring by using modredundant with the freeze function. The chair conformers were allowed to optimize without any restraints, since the high-energy conformers would minimize to the nearest saddle point and move out of its original coordinates. In the sugar model for hydrolysis with a variation in C1-OMe bond length, the C1-O1 bond was restrained to prevent the bond optimizing to its equilibrium bond length.

3.9.1 EQUATORIAL CHAIR SCANS

The equatorial chair conformer was allowed to optimize without constraints on the ring for a C1-OMe bond length of 1.40 Å – 2.40 Å. At a C1-OMe bond distance ≥ 2.00 Å, the chair conformer puckered into a half chair and envelope. An exocyclic hydroxyl scan was computed for C1-OMe bond lengths ≥ 1.70 Å. Conformational changes occurred at a C1-OMe bond distance of 1.80 Å but not before. At a C1-OMe bond distance of 1.75 Å, there was no conformational change. Each of the scan points were investigated for hydroxyl orientation and it was noted that at certain hydroxyl orientations, certain conformers were favored across C1-OMe bond lengths of 2.00 – 2.40 Å.

3.9.3 POST SCF ANALYSIS

The post SCF analyses conducted included natural hybrid orbitals (NHO's), NBO's[207], Merz-Kollman charges and QTAIM for electron densities. This type of post SCF analysis is useful in understanding some of the chemical changes that occur when a six-membered ring deforms and changes conformation. The chemical properties are charge, bond distances, electron densities and orientations of HOMO and LUMO orbitals. Each type of analysis looks at a different facet of chemical change and together can provide a holistic view of ring deformation.

3.9.3.1 ELECTRON DENSITY

The electron density can be thought of as a cloud of negative charge that varies in density at different parts of the molecule and is key to the bonding and geometry of a molecule because the interactions that occur between electrons and nuclei hold the molecule together. The repulsive forces that occur between electrons oppose the attractive forces that hold the molecule together. In the equilibrium geometry of the molecule, these forces are balanced or minimized.[208]

Once the selected basis function has been used to geometrically optimize the molecular structure, the electron density is obtainable. The electron density or total molecular density can be derived from the molecular wave function Ψ, as it is just the wave function squared.[107] The molecular orbitals that compose the wave function for a molecular system provide detailed information about the electronic interactions within a molecule. Since the wave function is written as a sum of products of individual molecular orbitals, which in turn consist of atomic orbitals, and all the molecular orbitals together give rise to the electron density.[110] The Atoms in Molecules method can be used to investigate the electron density and how it changes over the atoms in a molecule, which is discussed later in the chapter. As discussed earlier, each atomic orbital (AO) is composed of STOs or GTOs. The AOs are used to provide information about the atomic contributions that result in each molecular orbital.

Mulliken population analysis[209] and Natural population analysis[210] are used to describe the way charge is generally distributed and provide an estimate of the charge in a molecule. It can only be generally and not exactly determined since there is no charge eigen operator that can be applied to the quantum mechanical wave function. In quantum chemical calculations using the Schrödinger equation (see equation 2.8), the energy of the system can be calculated by providing an initial guess for the energy, then applying an operator ($\hat{H}$) to the wave function and solving for Ψ. The wave function is then used to obtain a better estimate of the energy, once the energy obtained has converged, the wave function can be used to calculate an observable if an operator is applied to it. In order to calculate an observable such as charge an operator is required, currently there is no defined operator available for calculating charge. Since it is difficult to determine exactly where the electron density is on the molecule at any given point. There is not a definite way to obtain charges from the wave function. Several methods exist which partition the electron density in order to estimate the charge on each atom.[209]

3.9.3.2 ATOMIC CHARGES

The electrostatic interactions between atoms are of key importance in systems with predominantly polar atoms, since the electrostatics dominates the nonbonding energy. These interactions can be treated at different levels of accuracy in classical methods (MM and MD). However, the atomic-centered point-charge model is used in most instances. In this technique each atom is assigned a partial charge, which has to be parameterized into the force field, and only the charge-charge Coulomb interaction is used.[211]

Mulliken population analysis is one of the simplest methods used to calculate charges in quantum chemical calculations. This method distributes charge according to the occupancy of the atomic orbitals with no differentiation to atom type and electronegativity.[212] Even though charge assignment varies with basis set[213], Mulliken charges are still used. Other orbital-based methods have been developed to account for the downfalls of the Mulliken charges[209] such as Löwdin population analysis[214] and natural population analysis[210].

Another approach to determining the atomic charge has been to derive them from a least-square fit electrostatic potential.[215,216] Methods that utilize a potential-based fit are CHELP (charges from electrostatic potential)[217], CHELPG and the Merz-Kollman scheme[218,219]. They differ mainly in the selection of points used to formulate the electrostatic potential. The Merz-Kollman scheme will be discussed here in detail. The Merz-Kollman scheme was chosen out of all the methods mentioned because the calculation of charges is conformationally sensitive.[211] Since conformational change was a central quantity in this study, applying this scheme was most appropriate.

3.9.3.3 MERZ-KOLLMAN

The Merz-Kollman[218,219] charges are calculated by fitting the atomic charges to reproduce the electrostatic potential of the molecule. The electrostatic potential is formulated by constructing multiple grid points located on several layers around the molecule. The layers are calculated as an overlay of the van der Waals spheres around each atom. After the van der Waals layer there are three more layers, each layer has a default scaling value of 1.4, 1.6, and 1.8. Once the electrostatic potential is computed from the valid grid points on all four layers, the atomic charges are then derived which reproduce the electrostatic potential as close as possible. The charges are fitted with a constraint that the sum of all the atomic charges must equal the total charge of the molecule.[220]

3.9.3.4 ATOMS IN MOLECULES THEORY

The Atoms In Molecules (AIM) method of R. F. W. Bader[221] was developed based on the quantum mechanical hypothesis that every aspect of a molecular system is stored in the wave function. Information on the energies and corresponding geometries of a molecular system are accessible from the wave function, as well as the connectivity of atoms, electronic charge distribution (electron density) and bond distances of a molecule.[221] The AIM method analyses the topology of the electron density $\rho(\boldsymbol{r})$ as a function of three spatial coordinates, by partitioning the molecular volume into atomic subspaces.[82] The electron density, which is equal to the square of the wave function, is integrated over n-1 coordinates, as depicted in equation 3.17.

$$\rho(r_1) = \int \left|\Psi(r_2, r_3, ..., r_n)\right|^2 dr_2 dr_3 ... dr_n \quad (3.17)$$

Each topological feature of electron density i.e. points of high electron density (maximum), points of low electron density (minimum) where the derivative is zero has an associated critical point. A critical point is labeled by calculating values for ω and σ, where ω is the number of non-zero curvatures and σ is the sum of the algebraic signs of the curvatures, the calculation of which will not be discussed in this thesis. Basin attractors are minima of the electron density surface and contain three negative eigenvalues and are termed bond critical points (3, -1). A bond critical point is found between every pair of nuclei and is interpreted as being a bond between atoms. A (3, +1) critical point occurs inside a ring of bonded atoms. The last critical point (3, +3) is found in the interior of a cage. The critical point we will look at in this study is the (3, -1), bond critical point.[222,223] The bond critical point is a point of maximum electron density in the bond. The position of the bond critical point shows the atom to which the electron density is favored to the most. The extent of charge accumulation in the bond increases with the number of electron pair bonds, hence the greater the number of electron pair bonds, the greater the value of the bond critical point, ρ_b and the bond order. Providing information on double and triple bond between atoms.[224]

3.9.3.5 NATURAL BOND ORBITAL THEORY

Natural bond orbitals (NBOs)[225] are calculated as a function of natural atomic orbitals. Initially all the orbitals associated with an atom such as the core orbitals and lone-pairs are localized around that atom as natural atomic orbitals.[226,227] Thereafter, utilizing the basis set atomic orbitals of the atoms involved in the bond results in localized orbitals of those atoms in the bond.[107] These sequential transformations from the starting basis set to natural atomic orbitals (NAOs), natural hybrid orbitals (NHOs), natural bond orbitals (NBOs) and natural localized molecular orbitals (NLMOs) result in the formation of canonical molecular orbitals (CMOs).[225,228] Molecular orbitals are generally more delocalized and difficult to interpret, therefore NHOs and NBOs are used, since these orbitals are more localized around the atoms in the molecular structure and are easier to interpret.

$$AOs \rightarrow NAOs \rightarrow NHOs \rightarrow NBOs \rightarrow NLMOs \rightarrow CMOs \quad (3.18)$$

In this thesis the primary objective behind using the NBO analysis method was to study the orbitals involved in stabilization of the charge developing at C1 as the C1-O1 bond is lengthened. The HOMO and LUMO orbital energies for this process are also considered when the pyranose ring changes conformation.

3.9.3.6 NATURAL HYBRID ORBITALS (NHO'S)

Orbital hybridization is one of the most useful concepts in valence theory.[229] Every natural bonding orbital is composed of two valence hybrid orbitals, h_A and h_B on atoms A and B with its corresponding polarization coefficients shown in equation 3.19.[228]

$$\sigma_{AB} = c_A h_A + c_B h_B \quad (3.19)$$

The polarization coefficients c_A and c_B indicate to which atom the electrons in the bond are drawn towards the most. NHO's are composed of a linear combination of natural atomic orbitals (NAO's), since to get to NBO's one has to first calculate NAO's then NHO's.[227]

3.10 PROTEIN DATA BANK (PDB)

A database search was conducted through the PDB for glycosyl hydrolase with β-D-glucose (BGC) in the enzyme pocket as the search criterion, with no differentiation to class of enzyme; the accumulation of this data will be used as experimental data. Glycosyl hydrolase enzymes as explained in Chapter 1, are enzymes that catalyse the hydrolysis between disaccharide units or glycans. Each PDB structure of glycosyl hydrolase enzyme was then analyzed for the pucker conformation of each pyranose ring in the enzyme. Sorting with respect to BGC, then removing duplicated data from possible repeated PDB structures, refined the data further. Each conformer of BGC obtained was stored in a list. The first occurrence of each conformer was selected and used to count the number of times that particular conformer occurred in the list. Initially the majority of the data had the $^4C_1$ conformer in the crystal structure. The

PDB structures were further analyzed by viewing each crystal structure in VMD, ensuring that each crystal structure was a glycosyl hydrolase enzyme. The journal article linked to the PDB structure was consulted. From the journal article, information about enzymatic activity and reaction mechanism was obtained and whether the β-D-glucose molecule was in the -1 position, which is the monosaccharide unit that undergoes hydrolysis and oxo-carbenium ion formation in the saccharide chain. The reaction mechanism provided information about which BGC pucker conformer was undergoing hydrolysis. From this information, one could decipher the particular conformer of BGC that was reacting.

CHAPTER 4

A STEREOELECTRONIC AND THERMODYNAMIC STUDY OF β-D-METHYL GLUCOSE CONFORMATIONAL CHANGES RELATED TO ANOMERIC CENTRE REACTIVITY

As discussed in Chapter 1, the hydrolysis of glycans occur through glycosidase enzymes utilizing an oxo-carbenium ion in the transition state.[3] The monosaccharide unit undergoing hydrolysis is described, in this thesis, in isolation of an enzyme binding pocket and incoming nucleophile with β-D-methyl glucose as the prototypical minimum motif. A variation in the C1-O1 bond length is investigated for β-D-methyl glucose, which is considered as the sugar model for hydrolysis.

4.1 POSSIBLE MECHANISMS IN HYDROLYSIS

As discussed in Section 1.6 during the hydrolysis of a glycosidic bond, the monosaccharide unit that is reacting undergoes a change in conformation. The enzymatic reaction for glycosyl hydrolase reactions, which is the focus of this thesis, results in the glycosidic bond being cleaved, in particular the C1-OR bond (R=monosaccharide unit or glycan). From Section 1.6 we note that in glycosidase reactions there is sometimes a mixture of S_N2 and S_N1[230] type character, as seen in phosphorylase and phosphoribosyltransferases enzyme reactions.[30-35] These reactions can be described as consisting out of A_N and D_N mechanisms.[231] For example, protein tyrosine phosphatase[232-234] utilizes an A_ND_N mechanism with the formation of a thiophosphate intermediate. The A_ND_N mechanism describes the approach of an electron-rich nucleophile towards the electron-deficient anomeric carbon, with the simultaneous retreat of the leaving group, when the leaving group is gradually ejected this refers to the A_N*D_N mechanism (see Figure 1.6).[28]

Glycosidase reactions also employ the D_NA_N and D_N*A_N mechanisms (Figure 4.1), these reactions are initiated by the dissociation of the leaving group, which is followed by the nucleophilic attack whilst the leaving group is still associated with

the anomeric carbon. This mechanism is known as D_NA_N. Alternatively, if the attack of the nucleophile occurs after the departure of the leaving group with the ability to attack from either face of the intermediate it is referred to as D_N*A_N mechanism as seen in Figure 4.1[28,235] where the asterisk indicates the formation of a short-lived intermediate that forms faster than it can diffuse away.

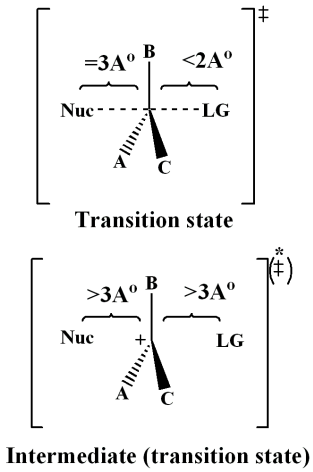


Figure 4.1 The D_NA_N and D_N*A_N mechanisms are characterized by a geometrically defined transition state or intermediate, with the predicted distances of the leaving group and nucleophile[28]

The mechanisms described above have a resultant positive charge placed on the anomeric carbon due to the dissociative nature of the transition state, to stabilize the carbocation the molecule can disperse electrons across the C1-O5 bond via hyperconjugation, with the possibility of forming an oxo-carbenium ion and reducing the positive charge on C1. If this occurs, electrons will be donated from the axial lone-pair electrons on the ring oxygen into the σ* anti-bonding orbital of the C1-O1 bond, which results in an increase of positive charge on the O5 atom since it is losing electrons.[26] The formation of the oxo-carbenium ion leads to an increase in planarity and double bond character of the C1-O5 bond and the molecule moving out of the $^4C_1$ conformer.[26] The hybridization on the C1 atom will move from sp^3 to sp^2 as the bond order across the C1-O5 bond increases. We will discuss the changes that occur during the glycosyl hydrolase D_N*A_N and D_NA_N reaction mechanism. The D_N*A_N mechanism is considered as a step-wise S_N1 mechanism since a short-lived oxo-

carbenium cation intermediate first forms which are subsequently attacked by a nucleophile. This thesis specifically considers the monosaccharide unit, which would be undergoing hydrolysis in the enzymatic binding pocket of the glycosyl hydrolase enzyme. Simulations are done in isolation from the protein and the incoming nucleophile, in order to understand the mechanism for conformational change of a monosaccharide unit during glycosidase reactions, from a chair to a half chair or envelope with lengthening of the bond between the anomeric carbon and the leaving group.

4.2 THE NATURE OF THE TRANSITION STATE

In many glycosyl hydrolase inverting and retaining reactions, the monosaccharide undergoing hydrolysis was found to be in a distorted conformer rather than the low energy $^{4}C_{1}$ or $^{1}C_{4}$ conformer.[26] Xevi Biarnés et al. have hypothesized that the monosaccharide unit would change into a conformer that allows for maximum oxo-carbenium ion formation.[15] These conformers are $^{4}H_{3}$, ^{4}E, $^{3}H_{4}$, ^{4}E, $B_{2,5}$, $^{2,5}B$ which place the C1-O5 in a planar position. These conformers in particular place the C2-C1-O5-C5 atoms in the same plane.[236,237] Davies and co-workers[26] states that if the chair conformer deforms into a twist boat for example $^{1}S_{3}$, then the twist boat conformer has to transition via a $^{4}H_{3}$ conformer, thereby also utilizing the oxo-carbenium ion transition state. The other hypothesis for a change in conformation of the monosaccharide unit is to reduce the 1,3 diaxial steric interactions between the hydrogen on the anomeric carbon and the incoming electron-rich nucleophile in the $^{1}S_{3}$ conformer for example.[26] The change in conformation also places the leaving group in a pseudoaxial position for a beta monosaccharide unit, which aids in the nucleophilic attack on the anomeric carbon.[15]

4.3 A SUGAR MODEL FOR HYDROLYSIS

In order to explain whether the reason for ring deformation in glycosyl hydrolase reactions is solely due to the formation of an oxo-carbenium ion a simple model compound, β-D-methyl glucose, was considered assuming that it is representative of all nine monosaccharide building blocks. This model compound is representative of a saccharide chain that would be present in a glycosyl hydrolase e.g. cellulase reaction which hydrolases β-1-4 linkages between glucose units.[238] To simplify the approach further only β-D-methyl glucose was considered and glycosyl hydrolase enzymes were the only enzymes studied. β-D-methyl glucose was considered since the methyl (Me) group attached to the anomeric carbon is a minimal motif that is representative of the extended saccharide R group attached to C1 in the enzymatic reaction.

The C1-O1 bond was gradually lengthened from 1.40 Å to 2.40 Å for β-D-methyl glucose. Free energy simulations were performed for β-D-methyl glucose with a variation in bond length of the C1-O1 bond. Coordinate data for all the conformers on the first hemisphere or latitude of the puckering conformational sphere[183] (E_5, $^4H_5$, ^4E, $^4H_3$, E_3, $^2H_3$, ^2E, $^2H_1$, E_1, $^OH_1$, ^OE and $^OH_5$) and the $^4C_1$ conformer were extracted and the charges on the C1, O5, O1 atoms and electron density across the C1-O5 bond were studied. Since for any glycosidase or glycosyl transferase reaction the substrate conformer is $^4C_1$. To change into any twist boat or boat conformer the chair conformer has to transition via one of these half chairs or envelopes. The D_NA_N mechanism considers the dissociation of the leaving group with the gradual approach of the nucleophile, in the transition state the leaving group and nucleophile are > 3.00 Å away from the anomeric carbon. In the transition state of the D_N*A_N[28] mechanism, the nucleophile is approximately 3.00 Å away from the anomeric carbon with the leaving group slightly dissociated at < 2.00 Å. The sugar model for hydrolysis in this thesis can therefore be considered to take into account both of these mechanisms. Since in both these mechanisms the nucleophile is a considerable distance away from the anomeric carbon.

The free energy simulations were done utilizing the SCC-DFTB Hamiltonian for molecular dynamics, of which the results were validated by calculations on the coordinate sets from FEARCF simulations with a B3LYP/6-311++G(d,p) basis set. Obtaining the global minima for each conformer of interest required an exhaustive hydroxyl conformational search and computing the potential energy at each point using both MM and SCC-DFTB parameters. Once the six low potential energy structures were obtained for each conformer, they were then optimized using B3LYP/6-311++G(d,p). The structure with the lowest energy from DFT was used for the C1-OMe bond stretch.

The triangular tessellation method[183] was used to quantitatively measure the change in ring deformation and assign each coordinate set obtained from the free energy simulations its corresponding conformer according to its triangular tessellation coordinates.[185] In order to sample all of phase space with respect to the triangular tessellation coordinates, one would have to bias the sampling during each iteration in the free energy simulation, forcing it to areas of phase space not yet sampled. Once all of phase space has been adequately sampled one would then have convergence of the free energy. Information regarding which conformer is the lowest in free energy as well as the minimum pathway required to traverse between a $^4C_1$ conformer across the free energy volume to the $^1C_4$ conformer, can be obtained. The reaction coordinate is represented by three theta angles (θ_0, θ_1, θ_2), which are the pucker angles defined by the six atoms in the ring. θ_0, θ_1 and θ_2 are calculated between atoms C4 C5 O5, O5 C1 C2 and C1 C2 C3 respectively. The biasing forces that are calculated from the pmf are applied to the theta angles by projecting the forces onto the Cartesian coordinates of the corresponding atoms, biasing the simulation away from previously sampled areas. The sugar model for hydrolysis with the C1-OMe bond gradually lengthened in increments of 0.05 Å from 1.40 Å - 1.90 Å then in increments of 0.1 Å from 2.00 Å - 2.20 Å, which simulates the glycosyl hydrolase reaction with the OR bond being gradually lengthened until the bond is dissociated. Free energy volumes were obtained for each of the increments.

The chemical properties of $^4C_1$ and all the pucker conformers on the first latitude of the Hill Reilly sphere were studied. The global minimum for each conformer was obtained; these coordinates were used in lengthening of the C1-OMe bond performing a geometry optimization at each point using B3LYP/6-311++G(d,p) whilst keeping the three ring pucker dihedral angles restrained during the optimization and also allowing the conformer to optimize without any constraints. All the conformers studied besides the $^4C_1$ conformer were optimized with constraints on the ring, to prevent a change in conformation, since the structure was optimized to a minimum at each bond change. It is expected that the high-energy conformers that are allowed to optimize without restraints, will optimize to the closest saddle point. The resultant wave function after geometry optimization was used to calculate the electron density of all bonds in the molecule and these profiles were analyzed using QTAIM. The Atoms in Molecule[222] method was discussed in Chapter 3. During the optimization, NBO's were calculated to provide information on charge transfer between atomic orbitals in the molecule. Merz-Kollman charges were used to obtain the charge on the C1, O1 and O5 atoms using the same level of theory for all the conformers of interest (See Appendix B).

As discussed in Section 1.6, during hydrolysis, the reaction can proceed via an inverting or retaining mechanism. The inverting reaction utilizes a S_N2 oxo-carbenium ion like transition state whilst the retaining reaction utilizes an oxo-carbenium ion double displacement mechanism. Whether the enzyme is utilizing a retaining or inverting mechanism both transition states utilize the oxo-carbenium ion character. Extraction of the monosaccharide unit from the enzymatic pocket, allows one to understand the electronic changes that occur when the monosaccharide ring changes conformer and if the resultant pucker helps facilitate the enzymatic reaction. If the monosaccharide unit were to adopt an oxo-carbenium ion, we expect the bond length of the C1-O5 bond to change, to become shorter since there is a partial double bond that forms due to an increased electron density across that bond. The charge on the C1 atom should become less positive and the charge on the O5 should become less negative since electron density is moving from the lone-pairs on the ring oxygen into the σ* anti-bonding of the C1-O1 bond.

4.4 FREE ENERGY VOLUMES

The iterative sampling of the FEARCF[153] method allows for all of phase space with respect to the pucker coordinates to be accessed. Free energy volumes are the result of a three-dimensional reaction coordinate set for ring pucker. The free energy volumes were obtained from the pmf file. The pmf data with respect to theta angles from the triangular decomposition method were compared to the known dictionary set of triangular tessellation coordinates for each canonical conformation in order to identify each conformer.

4.4.1 FREE ENERGY VOLUME FOR β-D-METHYL GLUCOSE

The free energy volume shown below for β-D-methyl glucose has an equilibrated bond length of 1.40 Å for the C1-O1 bond in vacuum using SCC-DFTB. The sampling of a three-dimensional pucker phase space for β-D-methyl glucose shows that the $^4C_1$ conformer sampled predominantly, which can be seen from the isosurfaces[239] shown in blue on the free energy volume in Figure 4.2. The $^4C_1$ conformer is the most energetically favorable conformer for β-D-methyl glucose. At 3 kcal/mol there is sampling of some twist boats and boats along the equator of the surface, with the $^1S_5$ conformer being the lowest in energy of the conformers on the equator.

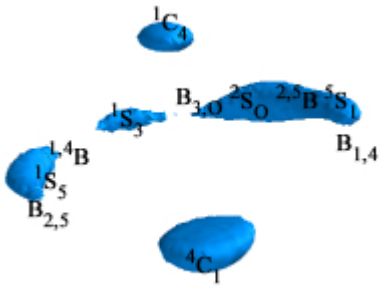


Figure 4.2 Free energy isosurface at 3 kcal/mol of β-D-methyl glucose with the C1-O1 bond equilibrated to 1.40 Å

The $^{1}S_{3}$ conformer in retaining β-Glycosidases is important since in both complexes of a GH20 chitobiase and a GH7 endoglucanase, the conformer was observed in the -1 subsite (reacting monosaccharide unit) and had transitioned via a ^{4}E conformer.[26] From the free energy volume, we note that the $^{4}H_{3}/^{4}E$ and $^{1}S_{3}$ conformers are in close approximation to one another, allowing for the pathway $^{4}C_{1} \rightarrow {}^{4}H_{3} \rightarrow {}^{1}S_{3}$.[240] The rationale for this change in pucker as mentioned before, is due to reduced 1,3 diaxial steric interaction between the incoming nucleophile and H1 and H3 atoms in the $^{1}S_{3}$ conformer. The presence of the ^{4}E conformer is consistent with the argument for oxo-carbenium ion formation. The free energy volume also shows that the lowest free energy half chair sampled from $^{4}C_{1}$ is the $^{4}H_{5}$ conformer, which is sampled at a free energy of 6.7 kcal/mol, which is not shown in the Figure 4.2 since the cut-off for the isosurfaces were set at 3 kcal/mol.

4.4.2 A MODEL FOR PUCKER CHANGE DURING HYDROLYSIS

Free energy simulations were performed on *β*-D-methyl glucose with a variation in bond length of the C1-OMe bond, from 1.55 Å to 2.20 Å and the free energy volumes are shown below (Figure 4.3). This illustrates a model for the change in bond length of the C1-O1 bond that occurs during hydrolysis of either D_NA_N or $D_N{*}A_N$ glycosidase reaction.

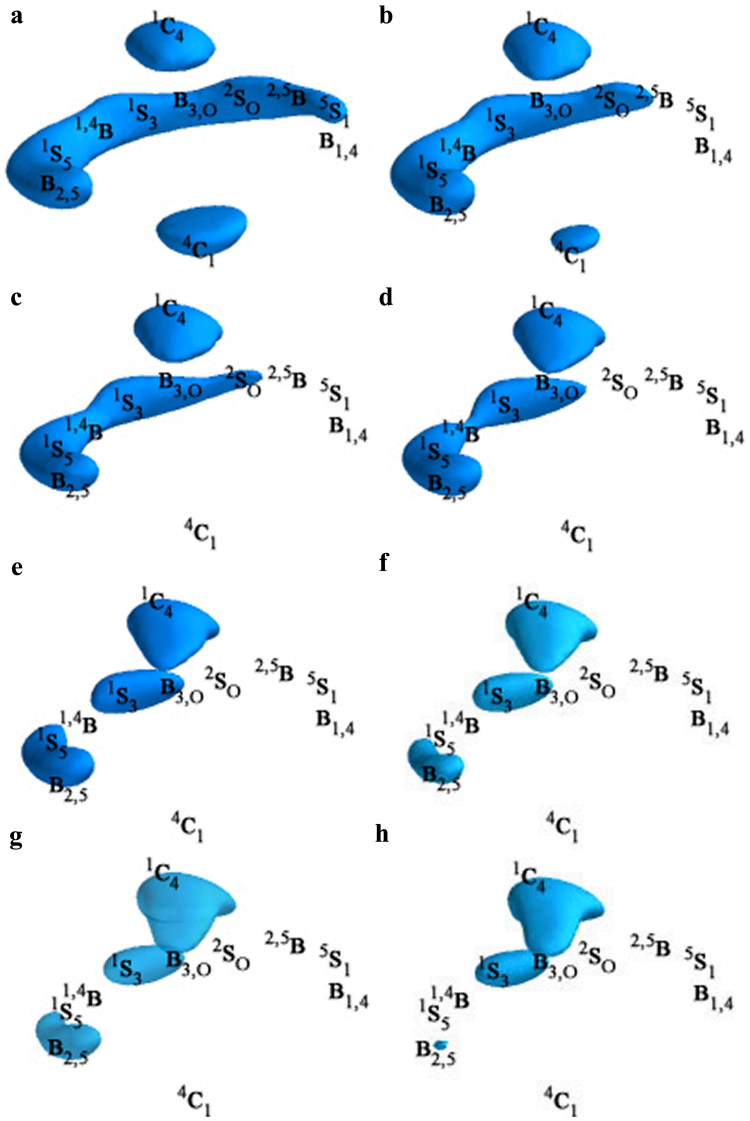


Figure 4.3 Free Energy Volumes of *β*-D-Methyl glucose with the C1-O1 bond constrained to a) 1.55 Å, b) 1.65 Å, c) 1.70 Å, d) 1.80 Å, e) 1.90 Å, f) 2.00 Å, g) 2.10 Å and h) 2.20 Å respectively

The free energy volume in Figure 4.3 shows that at an equilibrium bond length of the C1-O1 bond, the $^4C_1$ conformer is lower in free energy than the $^1C_4$ conformer as expected. But this does not hold true for when the bond length is increased until 2.20

Å, where the $^{1}C_{4}$ conformer is lower in free energy than the $^{4}C_{1}$ conformer. The $^{4}H_{5}$ and $^{4}C_{1}$ conformers become more energetically unfavorable as the C1-O1 bond length is increased, resulting in the sampling of these conformers at higher free energies. However the conformers used to cross from $^{4}C_{1}$ to the equator are the same as before, which leads to the minimum free energy pathways in Figure 4.4, with the $^{4}C_{1}$ conformer used as the reference.

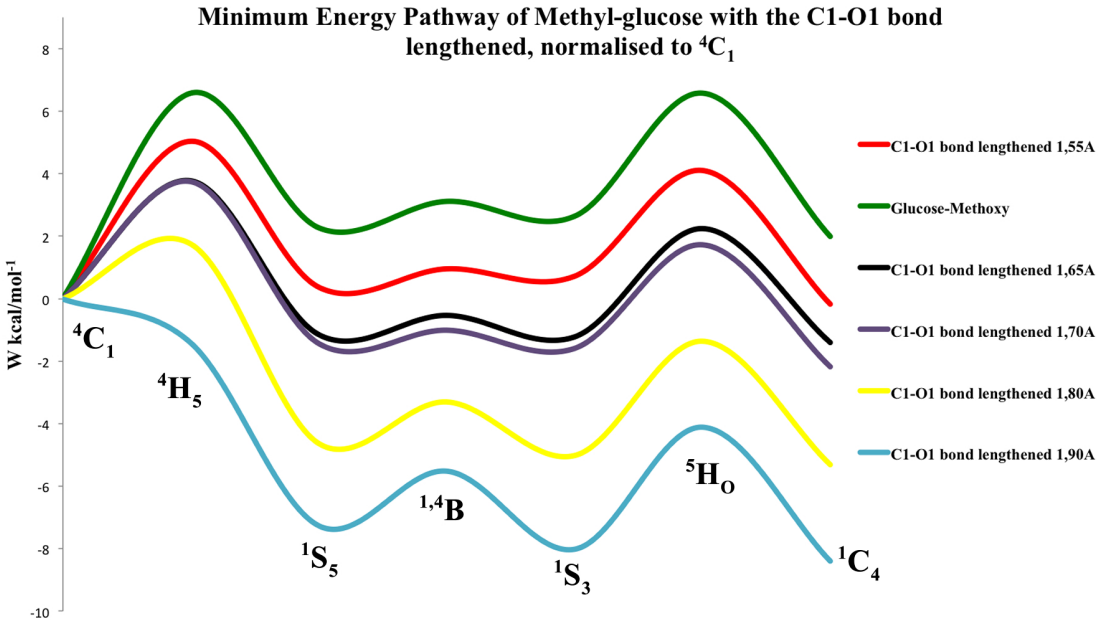


Figure 4.4 The minimum free energy pathway for *β*-D-methyl glucose with the C1-O1 bond at 1.40 Å, 1.55 Å, 1.65 Å, 1.70 Å, 1.80 Å, 1.90 Å, relative to the free energy of the $^{4}C_{1}$ conformer

A minimum free energy pathway is the lowest path between two points in free energy phase space. For ring pucker we often consider ring inversion pathway, a flip of the ring from a $^{4}C_{1}$ to $^{1}C_{4}$ conformer. The free energy values are zeroed at the $^{4}C_{1}$ conformer in order to compare energies of different free energy simulations. As the C1-O1 bond is lengthened the $^{4}C_{1}$ conformer becomes higher in free energy. The free energy of the $^{4}H_{5}$ conformer also increases but since the free energies are zeroed at $^{4}C_{1}$, the free energy then effectively decreases for the $^{4}H_{5}$ conformer as the C1-O1 bond is lengthened, with some conformers lower in energy than the $^{4}C_{1}$ conformer when the bond length is at 1.90 Å, suggesting that the $^{4}H_{5}$ conformer is a plausible transition state.

4.5 SELECTION OF TRANSITION STATES

From the free energy volumes (FEV), two-dimensional contours were plotted up to 15 kcal/mol through the first latitude of the pucker FEV (E_5, $^4H_5$, ^4E, $^4H_3$, E_3, $^2H_3$, ^2E, $^2H_1$, E_1, $^OH_1$, ^OE and $^OH_5$), for a C1-OMe bond distance of 1.40 Å – 2.20 Å. The contour slices were captured at iteration 17 and 8 for a C1-OMe bond distance of 1.40 Å - 1.90 Å and 2.00 Å - 2.20 Å respectively (Figure 4.5). The FEV at a C1-OMe equilibrium bond distance of 1.40 Å, showed the low energy transition state conformers to be $^4H_5$, ^4E, and E_5 with a free energy of approximately 6.7 kcal/mol. As the C1-OMe bond distance is lengthened, there is a tightening of the contours around the $^4H_5$ conformer. Indicating that this conformer is the lowest energy route out of the $^4C_1$ conformer, even as the C1-O1 bond lengthens. The blue spheres on the contours indicate the positions of each canonical conformer on the FEV. At C1-OMe bond lengths > 1.90 Å, show minimal sampling around conformers E_1, $^2H_1$, ^2E etc.

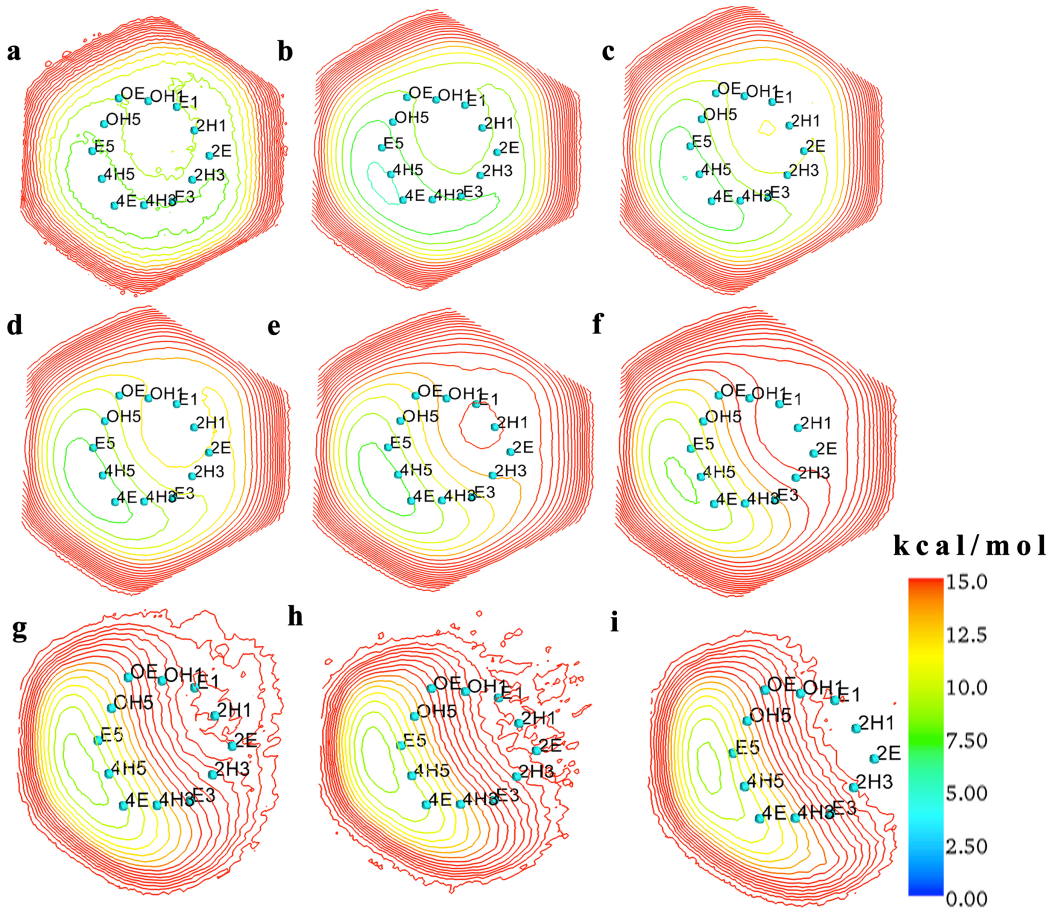


Figure 4.5 Contours through the first latitude in the FEV's at C1-O1 bond lengths of a) 1.40 Å, b) 1.55 Å, c) 1.65 Å, d) 1.70 Å, e) 1.80 Å, f) 1.90 Å, g) 2.00 Å, h) 2.10 Å and i) 2.20 Å respectively for β-D-methyl glucose

The transition states predicted by the free energy are the E_5, $^4H_5$ and ^4E conformers. From the FEV, canonical conformers on the first latitude through the pucker FEV were extracted and the global minimum obtained. Bond scans were conducted for a C1-OMe bond from 1.40 Å – 2.40 Å, with constraints on the ring dihedral angles for all conformers besides $^4C_1$. The corresponding DFT energies in kcal/mol are plotted in Figure 4.6.

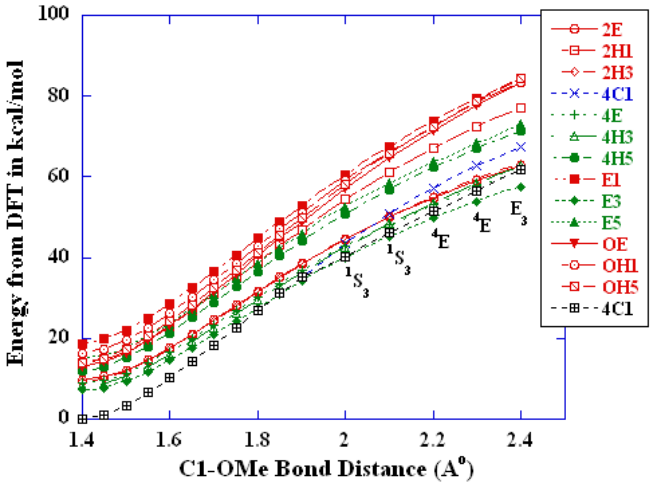


Figure 4.6 The change in energy for each conformer as the C1-OMe bond is lengthened (all the energies are zeroed to the equilibrium energy value for $^4C_1$ at a bond length of 1.40 Å)

The global minimum orientation of hydroxyls for each conformer on the first latitude through the pucker FEV (Figure 4.6) including $^4C_1$ was constant for each C1-OMe bond distance lengthened, in order to ensure that the change in the C1-OMe bond distance was the only variable affecting the energies, charges and electron densities. The red and green curves correspond to the high and low energy structures predicted by the FEV's. The low energy transition state conformers predicted by DFT are the ^4E, $^4H_3$ and E_3 with the $^4H_5$ and E_5 conformers higher in energy. The black curve corresponds to the chair conformer being optimized without any constraints for a C1-OMe bond length of 1.40 Å – 2.40 Å. At a C1-OMe bond length of 2.00 Å, there was a conformational change and the chair conformer optimized into a $^4H_3/^1S_3$ conformer. The representation is $^4H_3/^1S_3$ since the conformer is not entirely at the $^4H_3$ or $^1S_3$ canonical position, but inbetween. The chair conformer then optimizes to a ^4E conformer at a C1-OMe bond length of 2.20 Å and then an E_3 conformer at a C1-OMe bond length of 2.40 Å. The blue curve corresponds to the chair conformer constrained into a $^4C_1$ conformer, at a C1-OMe bond length ≥ 2.00 Å. The energy of the constrained $^4C_1$ conformer at this bond length is higher in energy than the unconstrained $^4C_1$ conformer that was allowed to optimize from a $^4C_1$ conformer into a $^1S_3$, ^4E and E_3 conformer. Indicating that at or close to the transition state as

predicted by DFT, the $^4C_1$ conformer is higher in energy than a puckered conformer such as $^4H_3$, ^4E or E_3.

4.6 MECHANISTIC PATHWAY FOR PUCKER

The mechanistic pathway from $^4C_1$ into a half chair or envelope as predicted by the free energy profile was the $^4H_5/^4E$ conformer. But the pathway predicted by DFT was the $^4E/^4H_3/E_3$ conformer.

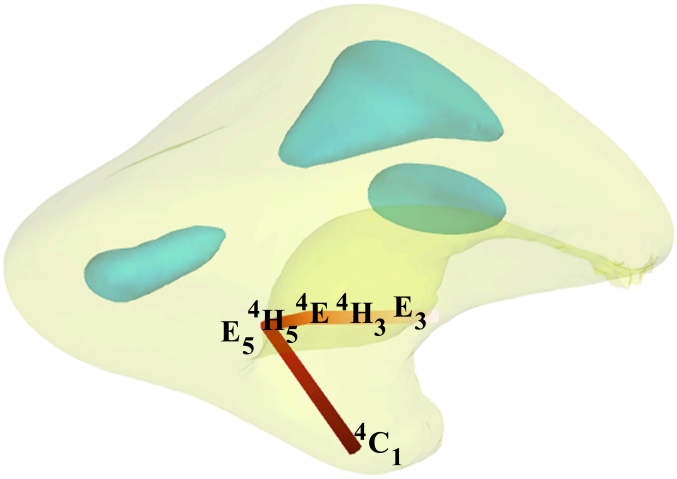


Figure 4.7 The hypothesized mechanistic pathway for pucker in glycosidase reactions involving *β*-D-methyl glucose, with the inner and outer isosurfaces plotted at 3 kcal/mol and 15 kcal/mol at a C1-OMe bond length of 2.00 Å

There is compatibility between the predictions of transition state pathways from free energy and density functional theory. Since the pathway predicted by the free energy is $^4C_1 \rightarrow {}^4H_5 \rightarrow {}^4E$, the pathway predicted by DFT shares the ^4E conformer, suggesting that a plausible mechanistic pathway could be $^4C_1 \rightarrow {}^4H_5 \rightarrow {}^4E \rightarrow {}^4H_3 \rightarrow E_3$, which is shown in Figure 4.7.

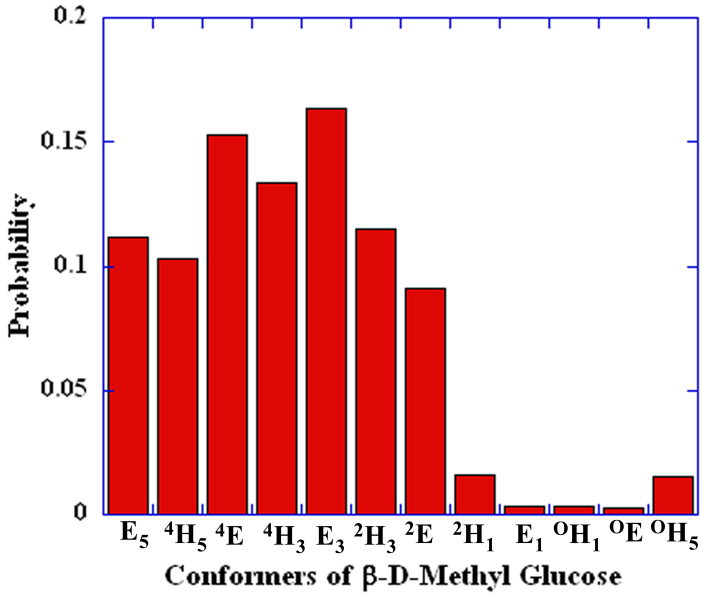
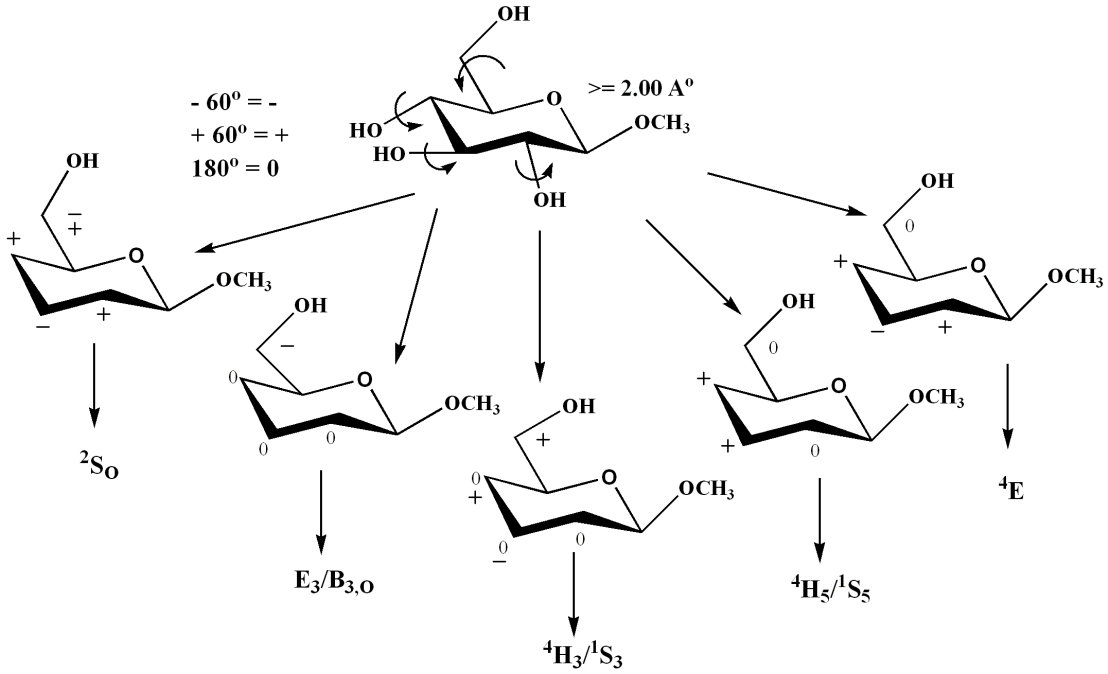


Figure 4.8 A statistical analysis of the combined trajectories until iteration six of the free energy simulation at a C1-OMe bond length of 2.00 Å

The relative statistical distributions for each conformer on the first latitude through pucker phase space were investigated. The group of conformers that are important as predicted by the free energy and DFT calculations are E_5, $^4H_5$, ^4E, $^4H_3$ and E_3, which are the predominantly sampled conformers (see Figure 4.8) and are predicted to be lower in energy according to DFT calculations and the FEV.

It can be postulated that the $^4H_5$ conformer is the conformer utilized to access the first latitude, the conformer then changes into ^4E, $^4H_3$ and E_3 at or close to the transition state in order to maximize orbital overlap and stabilizing intermolecular forces (e.g. hydrogen bonding). This type of conformational change was also visible in the mechanism of cellulose hydrolysis by Knott.[241] At the initial stage of the reaction, the -1 glucopyranose ring adopts the $^4H_5$ conformer. In the transition state the conformation of the glucopyranose ring in the -1 position changes to a $^4H_3$ conformer.[241] The ring deformation in β-1-4 cellulose hydrolysis reaction was also investigated in the scientific computing research unit laboratories and the mechanistic pathway of $^4C_1 \rightarrow {}^4H_5 \rightarrow {}^4E \rightarrow {}^4H_3 \rightarrow E_3$ was also shown to exist.[78]

The exocyclic hydroxyl orientation was varied for β-D-methyl glucose in the $^4C_1$ conformer at a C1-OMe bond length of 2.00 Å - 2.40 Å. DFT scans were computed from a C1-OMe bond distance of 2.00 Å since a conformational change occurred at this bond length. Geometry optimizations were conducted for each of the different hydroxyl orientations while constraining the breaking bond at 2.00 – 2.40 Å. At a particular combination of hydroxyl orientations a different conformer was favored. The change in orientation of hydroxyls can assist in getting to a particular conformer. This might explain how enzymes can chose particular conformers, by reorienting the hydroxyls to favor one conformer over another but does not explain why they pucker. An angle of -60 ^{o} (*gg*), +60 ^{o}(*gt*) and +180 ^{o}(*tg*) were used since these are the three low energy rotamers shown in Scheme 4.1.



Scheme 4.1 The orientation of each exocyclic hydroxyl angle as a function of the C1-OMe bond length constrained at ≥ 2.00 Å for the unconstrained $^4C_1$ conformer and it's corresponding optimized conformer as a result of the DFT scan

The double signs on some hydroxyl bond torsion angles indicate more than one accepted orientation for a specific angle; with that combination of hydroxyls and the same conformer will be obtained. The coordinates at each scan point were optimized to a minimum, visibility of twist boats are therefore expected.

4.7 STEREOELECTRONIC RATIONALE

The hypothesis proposed for ring deformation in glycosidase reaction by Xevi Biarnés et al.[15,236,237] has been discussed. The hypothesis for ring deformation by Davies and co-workers[26] suggested that a change in conformation of the monosaccharide unit was to reduce steric interactions between the hydrogen (H1) on the anomeric carbon and the incoming nucleophile.[26] The change in conformation also places the leaving group in a pseudoaxial position for a beta monosaccharide unit, which aids in the nucleophilic attack on the anomeric carbon by placing the H1 atom pseudoequatorial.[15]

The β-D-methyl glucose oxo-carbenium ion requires that the torsion angle τ_5 = C5-O5-C1-C2 becomes planar with a value of 0 ^{o}. When changing the bond length of the C1-OMe bond from 2.00 Å to 2.40 Å, a conformational change is noted for the unconstrained $^{4}C_{1}$ conformer, the dihedral angle changes around 1.90 Å from -58.34^{o} to -12.66^{o} as shown in Figure 4.9. The C5-O5-C1-H1 becomes planar at a C1-OMe bond length of > 1.90 Å, Scheme 4.2 shows the structural change in orientation of the C1-H1 bond. The C1-H1 bond becomes more planar closer to the transition state as the leaving group moves further away, there is a conformational change, to ensure less steric crowding for the incoming nucleophile, which is in accordance with the hypothesis proposed by Davies.[26]

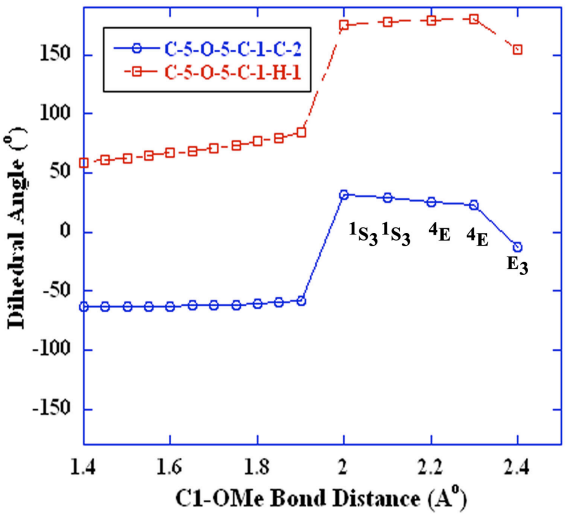
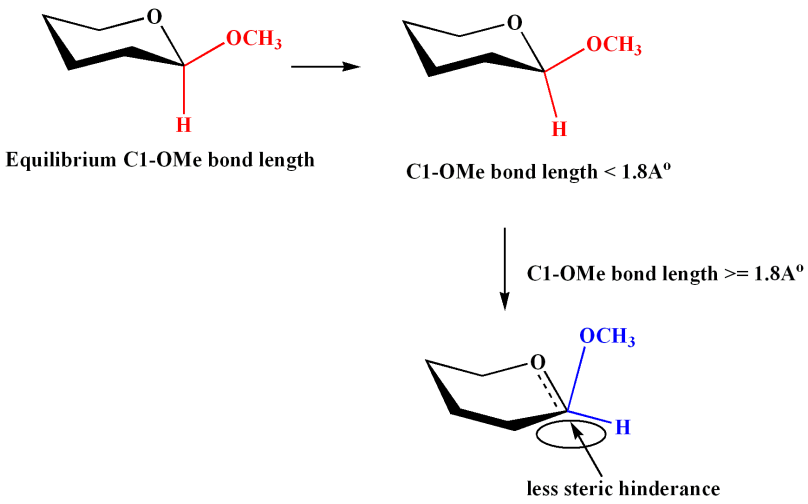


Figure 4.9 The C5-O5-C1-C2 and C5-O5-C1-H1 dihedral angles as a function of the C1-OMe bond length for the unconstrained $^4C_1$ conformer

The C5-O5-C1-C2 dihedral angle shown in Figure 4.9 changes from -58.34° to above 0° then decreases to -12.66°. The puckered conformer places the C5-O5-C1-C2 dihedral angle closer to planar than the chair conformer. The C5-O5-C1-H1 dihedral angle changes from 83.96° to 179.99°, with a gradual movement towards planarity.



Scheme 4.2 The conformational change of the C5-O5-C1-H1 torsion as the C1-OMe bond is lengthened from 1.40 Å to 2.40 Å

The gradual movement of the C5-O5-C1-H1 towards planarity as the C1-OMe bond is lengthened and as the chair conformer undergoes conformational change at a C1-O1 bond length of 2.00 Å with the global orientation of hydroxyls is shown in Scheme 4.2. The $^4C_1$ conformer also undergoes a conformational change at a C1-O1 bond distance of 1.80 Å since a DFT hydroxyl scan was conducted at this bond length and the conformation changed with certain combinations of hydroxyl orientations. A pseudoequatorial hydrogen allows for the approach of the nucleophile towards the anomeric carbon with minimal steric interactions in the $D_N{*}A_N$ mechanism, since the nucleophile does not approach the anomeric carbon until the leaving group is completely dissociated.

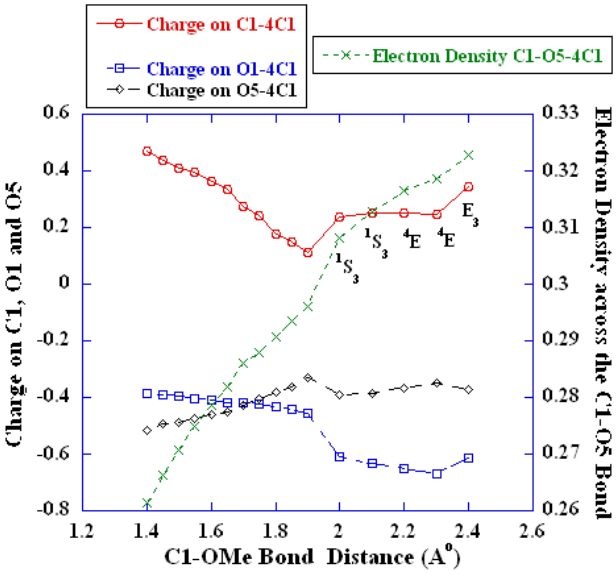


Figure 4.10 The charge on C1, O1, O5 and electron density across the C1-O5 bond as a function of the C1-OMe bond length for the unconstrained $^4C_1$ optimized conformer

The charge on C1, O5 and O1 and the electron density across the C1-O5 bond was recorded for the $^4C_1$ conformer with a conformational change at 2.00 Å. The charge on C1 becomes progressively less positive as the C1-OMe bond length is lengthened to 1.90 Å as shown in Figure 4.10. The electron density at the bond critical point across the C1-O5 bond increases for all bond lengths of the C1-O1 bond, there is however a greater increase in electron density when a conformational change occurs at a C1-OMe bond length of 2.00 Å. At a C1-OMe bond length of 2.00 Å, the charge

on C1 becomes more positive. Indicating that when the saccharide remains in a $^4C_1$ conformer, the anomeric carbon becomes progressively unattractive towards a nucleophile, as the leaving group dissociates. At the transition state the saccharide unit is required to pucker in order to create a slight positive charge on C1 to attract the nucleophile since in both the D_NA_N and $D_N{*}A_N$ mechanism the nucleophile is a considerable distance away from the anomeric carbon. All of the properties of an oxo-carbenium ion such as, the increase in electron density across the C1-O5 bond, the increase in bond order of the C1-O5 bond, the decrease in positive charge on C1 and the increase in positive charge on O5, are visible in the $^4C_1$ conformer. The charge on C1 becomes less positive and the charge on O5 becomes more positive since there is donation of lone-pair electrons from the ring oxygen into the σ* anti-bonding of the C1-O1 bond, which is depicted later in the chapter in Table 4.1.

A conformational change can indicate that during a glycosidase reaction, the change in pucker could favor the reaction mechanism. Since at 2.00 Å there is an increase in positive charge on the anomeric carbon, the O1 atom becomes more negative indicating that electron sharing between C1 and O1 has ceased and heterolytic cleavage occurs. The electron density across the C1-O1 bond decreases progressively as the bond is lengthened, which is expected since that bond is being broken (See Appendix A).

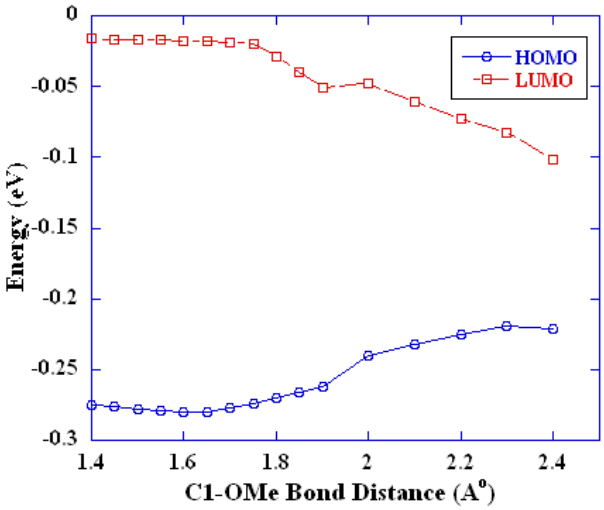


Figure 4.11 The energies for the HOMO and LUMO orbitals as a function of the C1-OMe bond length for the unconstrained $^4C_1$ conformer

The HOMO and LUMO energies for the unconstrained $^4C_1$ conformer as a function of the C1-OMe bond length are shown in Figure 4.11. These molecular orbital energies show relative small changes in energy for a C1-OMe bond length of 1.40 Å - 1.80 Å. At a conformational change, which occurs at 2.00 Å, there is a variation in energy. The LUMO orbital becomes more stable and lower in energy whilst the HOMO becomes higher in energy. The LUMO is the orbital, which the incoming nucleophile will attack, since this orbital is lower in energy, it would allow for a more facilitated, energetically favored attack of the anomeric carbon by the nucleophile. In terms of the glycosidase reaction, the HOMO can be thought of as the incoming nucleophile, in our model for hydrolysis however there is no nucleophile present. The leaving group can be interpreted as the nucleophile since it can reattach itself to the anomeric carbon.

Table 4.1 The donor NBO into the acceptor NBO and their corresponding delocalization stabilization energies as a function of the C1-OMe bond length for the unconstrained $^4C_1$ conformer with a conformational change at 2.00 Å

	C1-O1 (BD*)		**C1-O5 (BD*)**	
	O5 LP (1)	O5 LP (2)	O1 LP (1)	O1 LP (2)
C1-OMe bond distance	**Energy kcal/mol**			
1.40	3.93	-		13.31
1.45	4.11	-		10.80
1.50	4.28	-		8.81
1.55	4.47	-		7.21
1.60	4.66	-		5.94
1.65	4.84	-		4.91
1.70	5.00	-		4.07
1.75	5.04	0.68		3.38
1.80	4.83	1.76		2.82
1.85	4.26	3.53		2.35
1.90	3.42	5.73		1.97
2.00	-	48.50		1.38
2.10	-	54.71		0.95
2.20	-	60.39		0.65
2.30	-	65.51		-
2.40	-	66.59		0.51

NBO delocalization stabilization energies were calculated from an NBO analysis. Significant donations from a donor NBO into an acceptor NBO were noted for the lone-pairs on the ring oxygen (O5) donating into the antibonding orbital on the C1-O1 bond and the lone-pairs on the exocyclic oxygen (O1) donating into the antibonding orbital on the C1-O5 bond (Table 4.1).

Donation of lone-pair electrons on the exocyclic oxygen (O1) into the σ* anti-bonding orbital on the C1-O5 bond occurs. This type of electron donation from the exocyclic oxygen into the antibonding orbital on the C1-O5 bond is an example of the exo-anomeric effect. This orbital stabilization energy only occurs when the R group is gauche to the O5 atom. This is true for the unconstrained chair conformers investigated (Figure 4.12) before conformational change, where the R group is *gauche-trans* to O5 maximizing the gauche orbital interactions and minimizing steric interactions (Figure 1.7, *gt*). Initially this stabilization effect is important (13.31 kcal/mol) but it diminishes to almost 1 kcal/mol as the C1-O1 bond elongates. There is no increase in stabilization energy when donation of lone-pair electrons on the exocyclic oxygen into the σ* anti-bonding orbital on the C1-O5 bond occurs even at a C1-OMe bond distance of 2.00 Å where conformational change occurs. According to these calculations, the exo-anomeric effect plays very little role as the glycosidic bond lengthens into the transition state where the pyranose ring is puckered. This is expected in the context of this study.

At different bond lengths of the C1-OMe, there are donations from lone-pairs on the ring oxygen (O5) into the antibonding orbital shown in Table 4.1. At a conformational change, which occurs at 2.00 Å, there is an increase in stabilization energy from the donation of the lone-pair electrons on the O5 atom into the σ* anti-bonding orbital on the C1-O1 bond. This type of electron donation with the increased stabilization energy indicates that the anomeric effect is present (see Section 1.8.2). In Scheme 4.2 and Figure 4.12, note that as the C1-O1 bond lengthens the leaving group proceeds from being equatorial to pseudo-axial to axial, this is in line with the stabilizing stereoelectronic interactions predicted by the anomeric effect. Lone-pair two on the O5 atom (O5 LP (2)) donates into the σ* antibonding orbital on the C1-O1 bond at a bond distance of 1.75 Å with minimal stabilization energy, distances before that do not show donation from this lone-pair. The donation of electrons from O5 LP (2) starts occurring at a C1-O1 bond distance of 1.75 Å probably due to the C1-O1

bond becoming pseudoaxial at that bond distance. As discussed in Section 1.8.2, the anomeric effect occurs when the C1-O1 bond is axial with a significant increase in stabilization energy.

There is a noticeable increase in stabilization energy when there is a conformational change at 2.00 Å, which increases to a C1-O1 bond length of 2.40 Å. This donation of lone-pair electrons on the O5 atom into the antibonding orbital of the C1-O1 bond, is energetically more favorable with a change in conformation, which makes the C5-O5-C1-C2 and C5-O5-C1-H1 bonds more planar when compared to the $^{4}C_{1}$ conformer as well as the C1-O1 bond axial. At any C1-O1 bond length there is still donation from the lone-pair electrons on the O5 atom into the C1-O1 antibonding orbital but the anomeric effect occurs when there is a conformational change, which places the C1-O1 bond in the axial position. Since this conformational change results in an increased stabilization energy of the molecule.

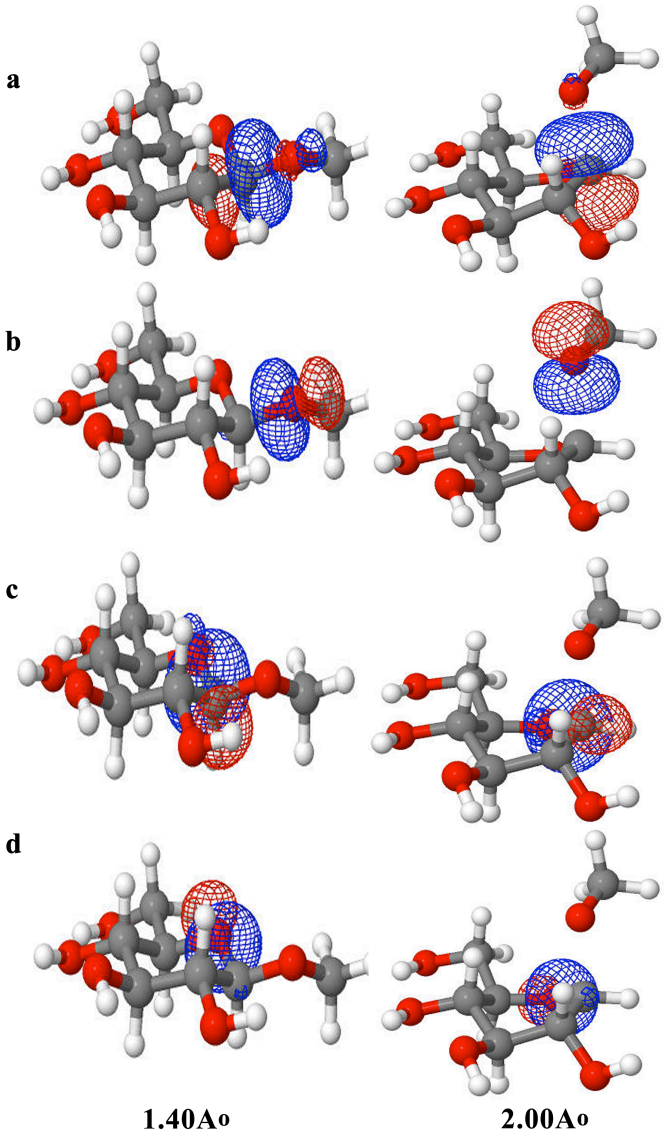


Figure 4.12 Natural Hybrid Orbitals (NHO) of a) C1 (O1), b) O1 (C1), c) C1 (O5) and d) O5 (C1) bonds for *β*-D-Methyl glucose with the C1-OMe bond length at 1.40 Å and 2.00 Å

The NHO orbitals for a C1-OMe bond length at 1.40 Å and 2.00 Å for bonds C1 (O1) and C1 (O5). The noticeable difference between these NHO's is the C1 (O1) bond, where in the $^{4}C_{1}$ conformer the LUMO (red lobe) is inaccessible to the incoming nucleophile because the approach of the nucleophile is sterically hindered by the H1 atom. At a C1-OMe bond length of 2.00 Å, where a conformational change occurs, the LUMO orbital is more accessible to the incoming nucleophile, since the H1 atom is now placed equatorial and there is reduced steric hindrance see Figure 4.12.

It can be hypothesized from the above results that a saccharide unit will be puckered, at or close to the transition state to increase the stabilization energy of the molecule by increasing the stabilization energy of the O5 lone-pair donation into the σ* anti-bonding orbital on the C1-O1 bond. The energy of the $^4C_1$ conformer with a C1-O1 bond distance of 2.00 Å, which is close to the transition state bond distance, is higher than the $^4C_1$ conformer that has changed conformation at the same C1-O1 bond distance see Figure 4.6. The chair conformer may also change conformation to lower the energy of the LUMO orbital, for a more energetically favorable nucleophilic attack on the anomeric carbon (Figure 4.11). The conformation also changes so that the C1 atom is sp^2-hybridized, to accommodate the increased electron density across the C1-O5 bond and the increased bond order from single bond to partial double bond. The C1 atom becomes sp^2-hybridized by making the C5-O5-C1-C2 bond planar, which occurs when there is a change in conformation (Figure 4.9). The C5-O5-C1-H1 dihedral angle also moves towards planarity when a conformational change occurs, allowing for a less sterically hindered approach of the nucleophile.

4.8 COMPARISON OF THEORETICAL AND EXPERIMENTAL DATA

The theoretical transition states obtained from FEARCF simulations and post SCF calculations were compared to the experimental data obtained from the Protein Data Bank (PDB).

4.8.1 PDB

A statistical analysis of the PDB was conducted for glycosyl hydrolase enzymes with BGC in the active site. Searching through the PDB with the keywords BGC and glycosyl hydrolase performed an initial crude analysis of the database. The results showed a predominance of $^4C_1$ conformers for glucose. Further refinements of this data with regards to conformers that are in the -1 position during the glycosidase reaction are shown in Figure 4.13.

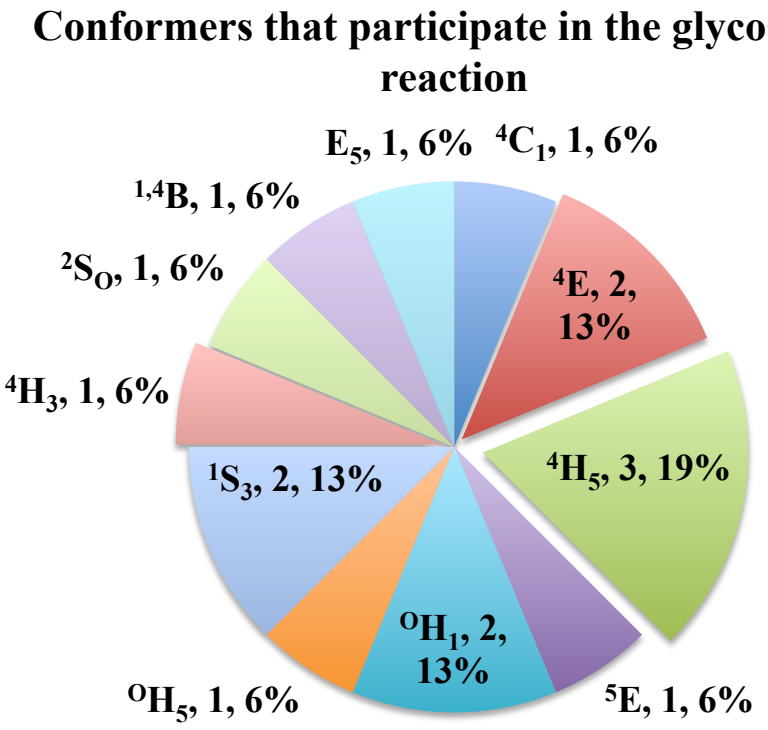


Figure 4.13 Conformers that are in the -1 position for glycosidase reactions

The group of conformers that are important as predicted by this study are E_5, $^4H_5$, ^4E, $^4H_3$ which appear in the PDB and consist of 44 % of all the conformers analyzed. A $^4C_1$ conformer was noted, as well as conformers $^OH_5$, and $^OH_1$, which make up 19 % of all of the conformers analyzed and for which there is no stereoelectronic rationale for at the moment. Conformers that are not on the first latitude of the pucker sphere but also appear in the PDB are $^1S_3$, $^{1,4}B$ and ^2S_O, which occur on the equator of pucker phase space. A change in conformation from $^4C_1$ to a twist boat or boat requires a transition via one of the half chairs or envelopes on the first latitude of pucker phase space. A $^1S_3$ conformer has to transition via a $^4H_3$ conformer, whilst a $^{1,4}B$ and ^2S_O conformer has to transition via conformer's ^4E and $^2H_3$ respectively and form 25 % of the conformers evaluated. The ^2S_O and $^2H_3$ conformers aren't on the direct mechanistic pathway observed in this study, probably due to these conformers appearing in inverting cellulases.[36,242,243] The $^4H_3$, ^4E conformers generally occur in retaining cellulase mechanisms.[244]

CHAPTER 5

CONCLUSION

The primary goal of this thesis was to investigate the rational for ring deformation of β-D-methyl glucose in cellulose hydrolysis, taking the variation in bond length of the C1-O1 bond as a model for hydrolysis. The investigation was computed in isolation from the enzyme binding pocket and incoming nucleophile. This objective has been achieved and some interesting results have been obtained.

The results for β-D-methyl glucose in gas phase with a C1-O1 bond stretch showed a transition path of $^4C_1 \rightarrow {}^4H_5 \rightarrow {}^4E \rightarrow {}^4H_3 \rightarrow E_3$. Previously the ring deformation was investigated in this laboratory and a similar mechanistic pathway was discovered for β-1-4 cellulose hydrolysis. Knott also discovered this mechanistic path.

Exocyclic hydroxyl analysis of β-D-methyl glucose in the unconstrained $^4C_1$ conformer with a C1-O1 bond length constrained at 2.00 – 2.40 Å showed different orientations and combination of hydroxyls with certain conformers being preferred. Torsional analysis of C5-O5-C1-C2 and C5-O5-C1-H1 angles for the $^4C_1$ conformer and when conformational change occurs showed a movement towards planarity of these dihedrals. The positive charge on the anomeric carbon decreases in the $^4C_1$ conformation, when a conformational change occurs the positive charge at C1 increases, making it more susceptible to nucleophilic attack. The process of ring deformation showed an increase in the stabilization energy of the molecule by increasing the stabilizing energy of the O5 lone-pair donation into the σ^* anti-bonding orbital on the C1-O1 bond. NBO analysis reveals the LUMO orbital being sterically crowded in the $^4C_1$ conformation, during the process of ring deformation the LUMO orbital becomes less sterically crowded and more accessible.

The pyranose ring therefore changes conformation to increase the planarity of the C5-O5-C1-C2 and C5-O5-C1-H1 dihedral angles, increase the positive charge on the anomeric carbon, and remove the steric hindrance around the anomeric carbon.

Further studies using the FEARCF method could be done with the other eight common monosaccharide units and more complex saccharide molecules. These

results could be used to provide further insight into the chemical role of ring deformation in hydrolysis reactions.

REFERENCES

(1) Gleeson, P. In *Carbohydrate-Protein Interaction*; Clarke, A., Wilson, I., Eds.; Springer Berlin Heidelberg: 1988; Vol. 139, p 1.
(2) Stryer, L. *Biochemistry*; 3 ed.; W H Freeman, 1988.
(3) Barnett, C. B.; Naidoo, K. J. *Journal of Physical Chemistry B* **2010**, *114*, 17142.
(4) Clark, D.; Sokoloff, L. *Basic Neurochemistry: Molecular, Cellular and Medical Aspects* 6th ed., 1999.
(5) Varki, A. *Nature* **2007**, *446*, 1023.
(6) Varki, N. M.; Varki, A. *Laboratory Investigation* **2007**, *87*, 851.
(7) Peterson, L.; Ardevol, A.; Rovira, C.; Reilly, P. J. *Journal of American Chemical Society* **2010**, *132*, 8291.
(8) Varki, A.; Cummings, R.; Esko, J.; Freeze, H.; Hart, G.; Marth, J. *Essentials of Glycobiology*; Second ed.; Cold Spring Harbor Laboratory Press, 1999.
(9) Rahi, H. *Biochemical Education* **1991**, *19*, 29.
(10) Wilson, J. L. *Journal of Chemical Education* **1988**, *65*, 783.
(11) Smith, E. L.; Hill, R. L.; Lehman, I. R.; Lefkowitz, R. J.; Handler, P.; White, A. *Principles of Biochemistry: General Aspects*; Seventh ed.; McGraw-Hill: New York, 1983.
(12) Berg, J. M.; Tymoczko, J. L.; Stryer, L. *Biochemistry*; W. H. Freeman and Company: New York, 2002.
(13) Stoddart, J. F. *Stereochemistry of Carbohydrates*; John Wiley & Sons: United States of America, 1971.
(14) Clayden, J.; Greeves, N.; Warren, S.; Wothers, P. *Organic Chemistry*; Oxford University Press, 2001.
(15) Biarnés, X.; Ardèvol, A.; Planas, A.; Rovira, C.; Laio, A.; Parrinello, M. *Journal of the American Chemical Society* **2007**.
(16) Davies, G. S.; Withers, S. G. *In Comprehensive Biological Catalysis*; London Academic, 1998.
(17) Lairson, L. L.; Henrissat, B.; Davies, G. J.; Withers, S. G. *Annual Review of Biochemistry* **2008**, *77*, 521.
(18) Takayama, S.; Chung, S. J.; Igarashi, Y.; Ichikawa, Y.; Sepp, A.; Lechler, R. I.; Wu, J.; Hayashi, T.; Siuzdak, G.; Wong, C. H. *Bioorganic & medicinal chemistry* **1999**, *7*, 401.
(19) Henrissat, B.; Davies, G. J. *Current Opinion in Structural Biology* **1997**, *7*, 637.
(20) *International Union of Biochemistry and Molecular Biology: Enzyme Nomenclature. Recommendations.*; AcademicPress: San Diego, 1992.
(21) Gilkes, N. R.; Claeyssens, M.; Aebersold, R.; Henrissat, B.; Meinke, A.; Morrison, H. D.; Kilburn, D. G.; Warren, R. A. J.; Miller, R. C. *European Journal of Biochemistry* **1991**, *202*, 367.
(22) Henrissat, B.; Bairoch, A. *Biochemical Journal* **1993**, *233*, 781.
(23) Koshland, D. E. *Biological Reviews* **1953**, *28*, 416.
(24) Davies, G.; Henrissat, B. *Structure* **1995**, *3*, 853.
(25) Sinnott, M. L. *Chemical Reviews* **1990**, *90*, 1171.
(26) Davies, G. J.; Planas, A.; Rovira, C. *Accounts of Chemical Research* **2011**, *45*, 308.
(27) Lindhorst, T. K. *Essentials of Carbohydrate Chemistry and Biochemistry*; John Wiley & Sons, 2003.
(28) Schramm, V. L.; Shi, W. *Current Opinion in Structural Biology* **2001**, *11*, 657.
(29) Imyanitov, N. S. *Journal of General Chemistry* **1990**, *60*, 417.
(30) Fedorov, A.; Shi, W.; Kicska, G.; Fedorov, E.; Tyler, P. C.; Furneaux, R. H.; Hanson, J. C.; Gainsford, G. J.; Larese, J. Z.; Schramm, V. L.; Almo, S. C. *Biochemistry* **2001**, *40*, 853.
(31) Shi, W.; Basso, L. A.; Santos, D. S.; Tyler, P. C.; Furneaux, R. H.; Blanchard, J. S.; Almo, S. C.; Schramm, V. L. *Biochemistry* **2001**, *40* 8204.
(32) Shi, W.; Li, C. M.; Tyler, P. C.; Furneaux, R. H.; Cahill, S. M.; Girvin, M. E.; Grubmeyer, C.; Schramm, V. L.; Almo, S. C. *Biochemistry* **1999**, *38* 9872.
(33) Focia, P. J.; Craig, S. P.; Eakin, A. E. *Biochemistry* **1998**, *37* 17120.
(34) Heroux, A.; White, E. L.; Ross, L. J.; Juzin, A. P.; Borhani, D. W. *Structure Folding and Design* **2000**, *8* 1309.

(35) Heroux, A.; White, E. L.; Ross, L. J.; Davis, R. L.; Borhani, D. W. *Biochemistry* **1999**, *38*, 14495.
(36) Davies, G. J.; Ducros, V. M.; Varrot, A.; Zechel, D. L. *Biochemical Society transactions* **2003**, *31*, 523.
(37) Flint, J.; Bolam, D. N.; Nurizzo, D.; Taylor, E. J.; Williamson, M. P.; Walters, C.; Davies, G. J.; Gilbert, H. J. *The Journal of biological chemistry* **2005**, *280*, 23718.
(38) Money, V. A.; Smith, N. L.; Scaffidi, A.; Stick, R. V.; Gilbert, H. J.; Davies, G. J. *Angewandte Chemie International Edition* **2006**, *45*, 5136.
(39) Biarnes, X.; Nieto, J.; Planas, A.; Rovira, C. *The Journal of biological chemistry* **2006**, *281*, 1432.
(40) Nerinckx, W.; Desmet, T.; Claeyssens, M. *Arkivoc* **2006**, *part 13*, 90.
(41) Brady, J. W. *Solvation: Carbohydrates. In Encylcopedia of Computational Chemistry*; First ed.; John Wiley & Sons: Chichester, 1998.
(42) Brady, J. W. *Journal of American Chemical Society* **1989**, *111*, 5156.
(43) Woods, R. J.; Dwek, R. A.; Edge, C. J. *Journal of Physical Chemistry* **1995**, *99*, 3832.
(44) Naidoo, K. J.; Brady, J. W. *Journal of American Chemical Society* **1999**, *121* 2244.
(45) Kuttel, M.; Brady, J. W.; Naidoo, K. J. *Journal of Computational Chemistry* **2002**, *23*, 1236.
(46) Bizzarri, A. R. C., S. *Journal of Physical Chemistry B* **2002**, *106*, 6617.
(47) Tsui, V.; Case, D. A. *Biopolymers* **2000**, *56*, 275.
(48) Isaac, E. D.; Shukla, A.; Platzman, P. M.; Hamann, D. R.; Barbiellini, B.; Tulk, C. A. *Journal of Physics and Chemistry of Solids* **2000**, *61*, 403.
(49) Kirby, A. J. *Stereoelectronic Effects*; Oxford University Press: Oxford and New York, 1996.
(50) Kirby, A. J. *The anomeric effect and related stereoelectronic effects at oxygen (reactiviity and structure)*; Springer-Verlag, 1983.
(51) Tvaroska, I.; Carver, J. P. *Journal of Physical Chemistry* **1994**, *98*, 9477.
(52) Cramer, C. J.; Truhlar, D. G.; French, A. D. *Carbohydrate Research* **1997**, *298*, 1.
(53) Barnett, C. B.; Naidoo, K. J. *Journal of Physical Chemistry B* **2008**, *112*, 15450.
(54) Stoddart, J. F. *Stereochemistry of Carbohydrates*; John Wiley & Sons, 1971.
(55) Toukach, F. V.; Ananikov, V. P. *Chemical Society Reviews* **2013**, *42*, 8376.
(56) Listowsky, I.; Avigad, G.; Englard, S. *Journal of American Chemical Society* **1965**, *87*, 1765.
(57) Kato, T.; Kinoshita, T. *Analytical Biochemistry* **1980**, *106*, 238.
(58) Kennedy, J. F.; Williams, N. R. *Carbohydrate Chemistry*; Royal Society of Chemistry, 1981; Vol. 12.
(59) Duus, J. Ø.; Gotfredsen, C. H.; Bock, K. *Electronic Journal of Pathology and Histology* **2002**, *8*.
(60) Bartlett, R. J.; Watts, J. D.; Schleyer, P. R.; Jorgensen, W. L.; Schaefer, H. F.; Schreiner, P. R.; Thiel, W.; Glen, R. *Endocyclopedia of Computational Chemistry*; John Wiley & Sons, 1998.
(61) Weast, R. C. *CRC Handbook of Chemistry and Physics*; 55 ed.; CRC Press, 1974.
(62) Rao, A. K. *Glycobiology* **2012**.
(63) Car, R.; Parrinello, M. *Physical Review Letters* **1985**, *55*, 2471.
(64) Kuttel, M. M.; Naidoo, K. J. *Journal of Physical Chemistry B* **2005**, *109*, 7468.
(65) Mason, P. E.; Neilson, G. W.; Enderby, J. E.; Saboungi, M.-L.; Cuello, G.; Brady, J. W. *The Journal of chemical physics* **2006**, *125*, 224505.
(66) McNamara, J. P.; Muslim, A.-M.; Abdel-Aal, H.; Wang, H.; Mohr, M.; Hillier, I. H.; Bryce, R. A. *Chemical Physics Letters* **2004**, *394*, 429.
(67) Yu-Jen, J. C.; Naidoo, K. J. *Journal of Physical Chemistry B* **2003**, *107*, 9558.
(68) Davis, F.; Seamus, P. J. H. *Biosensors & Bioelectronics* **2007**, *22*, 1224.
(69) Lynd, L. R.; Weimer, P. J.; van Zyl, W.; Pretorious, I. S. *Microbiology and Molecular Biology Reviews* **2002**, *66*, 506.
(70) Boraston, A. B.; Bolman, D. N.; Gibert, H. J.; Davies, G. J. *Biochemical Journal* **2004**, *384*, 769.
(71) Ragauskas, A. J.; Williams, C. K.; Davison, B. H.; Britovsek, G.; Cairney, J.; Eckert, C. A.; Frederick, W. J.; Hallett, J. O.; Leak, D. J.; Liotta, C. L. *Science* **2006**, *311*, 484.

(72) Himmel, M. E.; Ding, S. Y.; Johnson, D. K.; Adney, W. S.; Nimlos, M. R.; Brady, J. W.; Foust, T. D. *Science* **2007**, *315*, 804.
(73) Beckham, G. T.; Matthews, J. F.; Bomble, Y. J.; Bu, L.; Adney, W. S.; Himmel, M. E.; Nimlos, M. R.; Crowley, M. F. *Journal of Physical Chemistry B* **2010**, *114*, 1447.
(74) Kraulis, P. J.; Clore, G. M.; Nilges, M.; Jones, T. A.; Pettersson, G.; Knowles, J.; Gronenborn, A. M. *Biochemistry* **1989**, *28*, 7421.
(75) Divne, C. S., J.; Reinikainen, T.; Ruohonen, L.; Pettersson, G.; Knowles, J. K. C.; Teeri, T. T.; Jones, T. A. *Science* **1994**, *265*, 524.
(76) Divne, C. S., J.; Teeri, T. T.; Jones, T. A. *Journal of Molecular Biology* **1998**, *275*, 309.
(77) Lehtio, J.; Sugiyama, J.; Gustavsson, M.; Fransson, L.; Linder, M.; Teeri, T. T. *Proceedings of the National Academy of Sciences of the United States of America* **2003**, *100*, 484.
(78) Barnett, C. B.; Wilkinson, K. A.; Naidoo, K. J. *Journal of American Chemical Society* **2010**, *132*, 12800.
(79) Reif, F. *Fundamentals of Statistical and Thermal Physics*; McGraw-Hill, 1965.
(80) Pathria, R. K. *Statistical Mechanics*; Butterworth-Heinemann, 1965.
(81) Lebowitz, J. L.; Penrose, O. *Physics Today* **1973**, *26*, 23.
(82) Jensen, F. *Introduction to Computational Chemistry*; Jon Wiley and Sons, 1999.
(83) Leach, A. R. *Molecular Modelling Principles and Applications* Second ed.; Prentice Hall: England, 2001.
(84) Hemmingsen, L.; Madsen, D. E.; Esbensen, A. L.; Olsen, L.; Engelsen, S. B. *Carbohydrate Research* **2004**, *339*, 937.
(85) Guy, R. H. G.; Graham, W. *Computational Chemistry*; Oxford University Press, 1995.
(86) Hinchliffe, A. *Chemical Modeling from Atoms to Liquids* John Wiley & Sons, 1999.
(87) Montgomery, J. A.; Holmgren, S. L.; Chandler, D. *Journal of Chemical Physics* **1980**, *73*, 3688.
(88) Leach, A. R. *Molecular Modeling Principles and Applications*; Addison Wesley Longman Limited, 1996.
(89) Mackerell, A. D. *Journal of Computational Chemistry* **2004**, *25*, 1584.
(90) Lii, J.-L.; Allinger, N. L. *Journal of Computational Chemistry* **1991**, *12*, 186.
(91) van Duin, A. C.; Dasgupta, S.; Lorant, F.; Goddard, W. A. *Journal of Physical Chemistry A* **2001**, *105*, 9396.
(92) Naidoo, K. J. *Sci. China Chem.* **2011**, *54*, 1962.
(93) Case, D. A.; Cheatham, T. E.; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. *Journal of Computational Chemistry* **2005**, *26*, 1668.
(94) Brooks, B. R.; Brooks, C. L.; MacKerell, A. D.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S. *Journal of Computational Chemistry* **2009**, *30*, 1545.
(95) Scott, W. R. P.; Hünenberger, P. H.; Tironi, I. G.; Mark, A. E.; Billeter, S. R.; Fennen, J.; Torda, A. E.; Huber, T.; Krüger, P.; van Gunsteren, W. F. *Journal of Physical Chemistry A* **1999**, *103*, 3596.
(96) Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I.; Mackerell, A. D. *Journal of Computational Chemistry* **2010**, *31*, 671.
(97) Guvench, O.; Greene, S. N.; Kamath, G.; Brady, J. W.; Venable, R. M.; Pastor, R. W.; MacKerell, A. D. *Journal of Computational Chemistry* **2008**, *29*, 2543.
(98) Eichenberger, A. P.; Allison, J. R.; Dolenc, J.; Geerke, D. P.; Horta, B. A. C.; Meier, K.; Oostenbrink, C.; Schmid, N.; Steiner, D.; Wang, D.; van Gunsteren, W. F. *Journal of Chemical Theory and Computation* **2011**, *7*, 3379.
(99) Woods, R. J. *Glycoconjugate Journal* **1998**, *15*, 209.
(100) Hermans, J. *Advances in protein chemistry* **2005**, *72*, 105.
(101) Schrödinger, E. *Annalen der Physik (IV. Folge)* **1926**, *79*, 361.
(102) Schrödinger, E. *Annalen der Physik (IV. Folge)* **1926**, *80*, 437.
(103) Schrödinger, E. *Annalen der Physik (IV. Folge)* **1926**, *81*, 109.
(104) Born, M.; Oppenheimer, J. R. *Annalen der Physik (IV. Folge)* **1927**, *84*, 457.

(105) Goodisman, J. *Contemporary Quantum Chemistry An Introduction*; Plenum Publishing Corporation, 1977.
(106) French, A. P.; Taylor, E. F. *An Introduction to Quantum Physics*; Norton, 1978.
(107) Cramer, C. J. *Essentials of Computational Chemistry Theories and Models*; Second Edition ed.; Jon Wiley & Sons, 1961.
(108) MacDonald, J. K. L. *Physical Review* **1933**, *43*, 830.
(109) Slater, J. C. *Physical Review* **1930**, *35*, 210.
(110) Clark, T.; Koch, R. *The Chemist's Electronic Book of Orbitals*; Springer Berlin Heidelberg, 1999.
(111) Hartree, D. R. *Mathematical Proceedings of the Cambridge Philosophical Society* **1928**, *24*, 89.
(112) Atkins, P. W. *Molecular Quantum Mechanics Parts I and II: An Introduction to Quantum Chemistry*; Oxford University Press, 1977; Vol. 1.
(113) Roothaan, C. C. J. *Reviews of Modern Physics* **1951**, *23*, 69.
(114) Levine, I. N. *Quantum Chemistry* 4th ed ed.; Prentice Hall: Englewood Cliffs, New Jersey, 1991.
(115) Slater, J. C. *Physical Review* **1929**, *34*, 1293
(116) Hartree, D. R. *Proceedings of the Cambridge Philosophical Society* **1928**, *24* 111.
(117) Roothaan, C. C. J. *Reviews of Modern Physics* **1951**, *23*, 69.
(118) Fischer, C. F. *Computer Physics Communications* **1987**, *43*, 355.
(119) Gill, P. M. W. In *Advances in Quantum Chemistry*; John, R. S., Michael, C. Z., Eds.; Academic Press: 1994; Vol. Volume 25, p 141.
(120) Lewars, E. G. *Computational Chemistry Introduction to the Theory and Applications of Molecular and Quantum Mechanics*; 1st ed.; Kluwer Academic Publishers, 2003.
(121) Foresman, J. B. *Exploring Chemistry with Electronic Structure Methods*; Second ed. Pittsburgh, 1995.
(122) Koch, W.; Holthausen, M. C. *A chemist's guide to density functional theory*; Wiley-VCH: Weinheim, 2000.
(123) Hohenberg, P.; Kohn, W. *Physical Review B* **1964**, *1138*, 136.
(124) Kohn, W.; Sham, L. J. *Physical Review A* **1965a**, *137*, 1697.
(125) Becke, A. D. *The Journal of Chemical Physics* **1993**, *98*, 1372.
(126) Perdew, J. P.; Ernzerhof, M.; Burke, K. *Journal of Chemical Physics* **1996**, *105*, 9982.
(127) Elstner, M. *Theoretical Chemistry Accounts* **2006**, *116*, 316.
(128) van der Vaar, A.; Merz, K. M. *Journal of American Chemical Society* **1999**, *121*, 9182.
(129) Korth, M.; Pitonak, M.; Rezac, J.; Hobza, P. *Journal of Chemical Theory and Computation* **2010**, *6*, 344.
(130) McNamara, J. P.; Muslim, A.-M.; Adbel-Aal, H.; Wang, H.; Mohr, M.; Hillier, I. H.; Bryce, R. A. *Chemical Physics Letters* **2004**, *394*, 429.
(131) French, A. D.; Brady, J. W. *Journal of American Chemical Society* **1990**, 1.
(132) Alkorta, I.; Popelier, P. L. A. *Carbohydrate Research* **2011**, *346*, 2933.
(133) Cumming, D. A.; Carver, J. P. *Biochemistry* **1987**, *26*, 6664.
(134) Bock, K. *Pure and Applied Chemistry* **1983**, *55*, 605.
(135) Thogersen, H.; Lemieux, R. U.; Bock, K.; Meyer, B. *Canadian Journal of Chemistry* **1982**, *79*, 2361.
(136) Allinger, N. L. *Journal of American Chemical Society* **1977**, *99*, 8127.
(137) Allinger, N. L.; Yuh, Y. H.; Lii, J.-H. *Journal of American Chemical Society* **1989**, *111*, 8551.
(138) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *Journal of American Chemical Society* **1995**, *117*, 5179.
(139) Momany, F. A.; Willett, J. L. *Carbohydrate Research* **2000**, *326*, 194.
(140) Kroon–Batenburg, L. M. J.; Kruiskamp, P. H.; Vliegenthart, J. F. G.; Kroon, J. *Journal of Physical Chemistry B* **1997**, *101*, 8454.
(141) Ott, K.-H.; Meyer, B. *Journal of Computational Chemistry* **1996**, *17*, 1068.
(142) Lins, R. D.; Hunenberger, P. H. *Journal of Computational Chemistry* **2005**, *26*, 1400.
(143) Pol-Fachin, L.; Rusu, V. H.; Verli, H.; Lins, R. D. *Journal of Chemical Theory and Computation* **2012**, *8*, 4681.

(144) Cramer, C. J.; Truhlar, D. G. *Journal of American Chemical Society* **1993**, *115*, 5745.
(145) Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I.; Mackerell, A. D. *Journal of Computational Chemistry* **2010**, *31*, 671.
(146) Raman, E. P.; Guvench, O.; MacKerell, A. D. *Journal of Physical Chemistry* **2010**, *114*, 12981.
(147) Guvench, O.; Hatcher, E.; Venable, R. M.; Pastor, R. W.; MacKerell, A. D. *Journal of Chemical Theory and Computation* **2009**, *5*, 2353.
(148) Guvench, O.; Mallajosyula, S. S.; Raman, E. P.; Hatcher, E.; Vanommeslaeghe, K.; Foster, T. J.; Jamison, F. W.; MacKerell, A. D. *Journal of Chemical Theory and Computation* **2011**, *7*, 3162.
(149) Hatcher, E. R.; Guvench, O.; MacKerell, A. D. *Journal of Chemical Theory and Computation* **2009**, *5*, 1315.
(150) Ragazzi, M.; Ferro, D. R.; Provasoli, A. *Journal of Computational Chemistry* **1986**, *7*, 105.
(151) Momany, F. A.; Willet, J. L. *Journal of Computational Chemistry* **2000**, *21*, 1204.
(152) Wales, D. J. *Science* **2001**, *293*, 2067.
(153) Naidoo, K. J. *Physical Chemistry Chemical Physics* **2012**, *14*, 9026.
(154) Ensing, B.; De Vivo, M.; Liu, Z. W.; Moore, P.; Klein, M. L. *Accounts of Chemical Research* **2006**, *39*, 73.
(155) Wolynes, P. G.; Onuchic, J. N.; Thirumalai, D. *Science* **1995**, *267*, 1619.
(156) Fersht, A. *Structure and mechanism in protein science: a guide to enzyme catalysis and protein folding*; W. H. Freeman, 1999.
(157) Karplus, M. *Nature Chemical Biology* **2011**, *17*, 401.
(158) Perrot, P. *A to Z of Thermodynamics* Oxford University Press, 1998.
(159) Zwanzig, R. W. *Journal of Chemical Physics* **1954**, *22*, 1420.
(160) Mezei, M.; Beveridge, D. L. *Annals | The New York Academy of Sciences* **1986**, *482* 1.
(161) Otter, W. K. D. *Journal of Chemical Theory and Computation* **2013**, *9*, 3861.
(162) Huber, T.; Torda, A. E.; van Gunsteren, W. F. *Journal of Computer-Aided Molecular Design* **1994**, *8*, 695.
(163) Cyr, E. C.; Bond, S. D. *Journal of Computational Physics* **2007**, *225*, 714.
(164) Kirkwood, J. G. *Journal of Chemical Physics* **1935**, *3*, 300.
(165) Shankar, K.; John, R. M.; Djamal, B.; Robert, H. S.; Peter, A. K. *Journal of Computational Chemistry* **1992**, *13* 1011.
(166) Jorgensen, W. L.; Madura, J. D.; Swenson, C. J. *Journal of American Chemical Society* **1984**, *106*, 6638.
(167) Hansen, H. S.; Hunenberger, P. H. *Journal of Chemical Theory and Computation* **2010**, *6*, 2622.
(168) Hansen, H. S.; Hunenberger, P. H. *Journal of Computational Chemistry* **2009**, *31*, 1.
(169) Darve, E.; Pohorille, A. *Journal of Chemical Physics* **2001**, *115* 9169.
(170) Darve, E.; Wilson, M. A.; Pohorille, A. *Molecular Simulation* **2002**, *28*, 113.
(171) Rodriguez-Gomez, D.; Darve, E.; Pohorille, A. *Journal of Chemical Physics* **2004**, *120*, 3563.
(172) He´nin, J.; Chipot, C. *Journal of Chemical Physics* **2004**, *121*, 2904.
(173) Rajamani, R.; Naidoo, K. J.; Gao, J. *Journal of Computational Chemistry* **2003**, *24*, 1775.
(174) Berg, B. A.; Neuhaus, T. *Physical Review Letters* **1992**, *68*, 9.
(175) Lee, J. *Physical Review Letters* **1993**, *71*, 211.
(176) Wang, F. G.; Landau, D. P. *Physical Review Letters* **2001**, *86*, 2050.
(177) F. G. Wang, D. P. L. *Physical Review E - Statistical, Nonlinear and Soft Matter Physics* **2001**, *64*, 56101.
(178) Barnett, C. B.; Naidoo, K. J. *Molecular Physics* **2009**, *107*, 1243.
(179) *Pure and Applied Chemistry* **1981**, *53*, 1901.
(180) Kilpatrick, J. E.; Pitzer, K. S.; Spitzer, R. *Journal of American Chemical Society* **1947**, *69*, 2483.
(181) Cremer, D.; Pople, J. A. *Journal of the American Chemical Society* **1975**, *97*, 1354.
(182) Sega, M.; Autieri, E.; Pederiva, F. *Journal of Chemical Physics* **2009**, *130*, 225102.

(183) Hill, A. D.; Reilly, P. J. *Journal of Chemical Information and Modeling* **2007**, *47*, 1031.
(184) Evans, D. G.; Boeyens, J. C. A. *Acta Crystallographica, Section B: Structural Science* **1989**, *45*, 581.
(185) Khalili, P.; Barnett, C. B.; Naidoo, K. J. *The Journal of Chemical Physics* **2013**, *138*, 184110.
(186) Ernest, D.; David, F. *Chemical Reviews (Washington, DC, United States)* **1986**, *86*, 681.
(187) Ditchfield, R.; Hehre, W. J.; Pople, J. A. *Journal of Physical Chemistry* **1971**, *54*, 724.
(188) Hehre, W. J.; Ditchfield, R.; Pople, J. A. *Journal of Physical Chemistry* **1972**, *56*, 2257.
(189) McLean, A. D.; Chandler, G. S. *Journal of Physical Chemistry* **1980**, *72*, 5639.
(190) Finch, P. *Carbohydrates: Structure, Syntheses and Dynamics*; Springer, 1999.
(191) Barrows, S. E.; Dulles, F. J.; Cramer, C. J.; French, A. D.; Truhlar, D. G. *Carbohydrate Research* **1995**, *276*, 219.
(192) Csonka, G. I.; French, A. D.; Johnson, G. P.; Stortz, C. A. *Journal of Chemical Theory and Computation* **2009**, *5*, 679.
(193) Kohn, W.; Becke, A. D.; Parr, R. G. *Journal of Physical Chemistry* **1996**, *100*, 12974.
(194) Csonka, G. I.; Elias, K.; Csizmadia, I. G. *Journal of Computational Chemistry* **1997**, *18*, 330.
(195) Csonka, G. I. *Journal of Molecular Structure* **2002**, *584*, 1.
(196) Csonka, G. I.; Elias, K.; Csizmadia, I. G. *Chemical Physics Letters* **1996**, *257*, 49.
(197) Lii, J.-H.; Ma, B.; Allinger, N. L. *Journal of Computational Chemistry* **1999**, *20*, 1593.
(198) Strati, G. L.; Willett, J. L.; Momany, F. A. *Carbohydrate Research* **2002**, *337*, 1833.
(199) Momany, F. A.; Appell, M.; Strati, G.; Willett, J. L. *Carbohydrate Research* **2004**, *339*, 553.
(200) Frank A. Momany, M. A., J. L. Willet, Wayne B. Bosma *Carbohydr. Res.* **2005**, *340*, 459.
(201) Becke, A. D. *Journal of Physical Chemistry* **1993**, *98*, 5648.
(202) Andre, I.; Tvaroska, I.; Carver, J. P. *Carbohydrate Research* **2003**, *338*, 865.
(203) Tvaroska, I.; Andre, I.; Carver, J. P. *Glycobiology* **2003**, *13*, 559.
(204) Brown, J. W. W., B. D. *Journal of American Chemical Society* **1996**, *118*, 1190.
(205) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *Journal of Computational Chemistry* **1983**, *4*, 187.
(206) Krishnan, R.; Binkley, J. S.; Seeger, R.; Pople, J. A. *Journal of Chemical Physics* **1980**, *72*, 650.
(207) Löwdin, P.-O. *Physics Review* **1955**, *97*, 1474.
(208) Matta, C. F.; Gillespie, R. J. *Journal of Chemical Education* **2002**, *79*, 1141.
(209) Bachrach, S. M. *Reviews in Computational Chemistry* **1994**, *5*, 171.
(210) Reed, A. E.; Weinstock, R. B.; Weinhold, F. A. *Journal of Chemical Physics* **1985**, *83*, 735.
(211) Sigfridsson, E.; Ryde, U. *Journal of Computational Chemistry* **1997**, *19*, 377.
(212) Mulliken, R. S. *Journal of Chemical Physics* **1962**, *36*, 3428.
(213) Mulliken, R. S.; Politzer, P. *Journal of Chemical Physics* **1971**, *55*, 5135.
(214) Lowdin, P.-O. *Advances in Quantum Chemistry* **1970**, *5*, 185.
(215) Chipot, C.; Maigret, B.; Rivail, J.-L.; Scheraga, H. A. *Journal of Physical Chemistry* **1992**, *96*, 10276.
(216) Angyan, J. G.; Chipot, C. *Chemical Physics Letters* **1995**, *241*, 51.
(217) Chirlain, L. E.; Francl, M. M. *Journal of Computational Chemistry* **1987**, *6*, 894.
(218) Singh, U. C.; Kollman, P. A. *Journal of Computational Chemistry* **1984**, *5* 129.
(219) Besler, B. H.; Merz., J. K. M.; Kollman, P. A. *Journal of Computational Chemistry* **1990**, *11* 431
(220) Ramachandran, K. I.; Deepa, G.; Namboori, K. *Computational Chemistry and Molecular Modeling: Principles and Applications* Springer, 2008.
(221) Bader, R. F. W. *Accounts of Chemical Research* **1985**, *18*, 9.
(222) Bader, R. F. W. *Atoms in Molecules - A Quantum Theory* Oxford University Press: Oxford, 1990.

(223) Bader, R. F. W. *Chemical Reviews* **1991**, *91*, 893.
(224) Bader, R. F. W. *Atoms in Molecules: A Quantum Theory*; Oxford University Press: USA, 1994.
(225) Reed, A. E.; Curtiss, L. A.; Weinhold, F. *Chemical Reviews (Washington, DC, United States)* **1988**, *88*, 899.
(226) Weinhold, F. *Natural Bond Orbital Methods*; John Wiley & Sons: Chichester, UK, 1998; Vol. 3.
(227) Weinhold, F. L., Clark R. *Discovering Chemistry With Natural Bond Orbitals*; John Wiley & Sons: New Jersey, 2012.
(228) Weinhold, F.; Landis, C. R. *Chemistry Education Research and Practice* **2001**, *2*, 91.
(229) Foster, J. P.; Weinhold, F. *Journal of American Chemical Society* **1980**, *102*, 7211.
(230) Vasella, A.; Davies, G. J.; Bohm, M. *Current Opinion in Chemical Biology* **2002**, *6*, 619.
(231) Guthrie, R. D.; Jencks, W. P. *Accounts of Chemical Research* **1989**, *22*, 343.
(232) Hoff, R. H.; Hengge, A. C.; Wu, L.; Keng, Y. F.; Zhang, Z. Y. *Biochemistry* **2000**, *39* 46.
(233) Zhang, M.; Zhou, M.; Van Etten, R. L.; Stauffacher, C. V. *Biochemistry* **1997**, *36*, 15.
(234) Fauman, E. B.; Yuvaniyama, C.; Schubert, H. L.; Stuckey, J. A.; Saper, M. A. *The Journal of biological chemistry* **1996**, *271* 18780.
(235) Guthrie, R. D. *Pure and Applied Chemistry* **1989**, *61*, 23.
(236) Davies, G. J.; Ducros, V. M.-A.; Varrot, A.; Zechel, D. L. *Biochemical Society transactions* **2003**, *31*, 523.
(237) Diego, G. M. A.; Marie-Bernard, L.; Marcelo, C.; Hélène, S.; Victor, L.; Pierre, B.; Pedro, A. M. *Journal of Molecular Biology* **2002**, *316*, 1061.
(238) Sukharnikov, L. O.; Alahuhta, M.; Brunecky, R.; Upadhyay, A.; Himmel, M. E.; Lunin, V. V.; Zhulin, I. B. *The Journal of biological chemistry* **2012**, *287*, 41068.
(239) Hansen, C. D.; Johnson, C. R. *Visualization Handbook*; Academic Press, 2004.
(240) Mayes, H. B.; Broadbelt, L. J.; Beckham, G. T. *Journal of American Chemical Society* **2014**, *136*, 1008.
(241) Knott, B. C.; Momeni, M. H.; Crowley, F. M.; Mackenzie, L. F.; Götz, A. W.; Sandgren, M.; Withers, S. G.; Ståhlberg, J.; Beckham, G. T. *Journal of American Chemical Society* **2014**, *136*, 321.
(242) Cantarel, B. L.; Coutinho, P. M.; Rancurel, C.; Bernard, T.; Lombard, V.; Henrissat, B. *Nucleic Acids Research* **2009**, *37*, D233.
(243) Fushinobu, S.; Mertz, B.; Hill, A. D.; Hidaka, M.; Kitaoka, M.; Reilly, P. J. *Carbohydrate Research* **2008**, *343*, 1023.
(244) Satoh, H.; Manabe, S. *Chemical Society Reviews* **2013**.

APPENDIX A

Tables of β-D-Methyl Glucose for conformers $^4C_1$, E_5, $^4H_5$, ^4E, $^4H_3$, E_3, $^2H_3$, ^2E, $^2H_1$, E_1, $^OH_1$, ^OE and $^OH_5$

A.1 Bond distances, electron densities and charges

Table A.1 C1-O1, C1-O5 bond distances, electron densities and charges on C1, O5, O1 for $^4C_1$, ^2E, $^2H_1$

CONFORMER	$^4C_1$		^2E		$^2H_1$	
C1-O1 BOND LENGTH						
1.40 Å	Bond Distance(Å)/Electron Density/Charge		Bond Distance(Å)/Electron Density/Charge		Bond Distance(Å)/Electron Density/Charge	
	C1-O1	C1-O5	C1-O1	C1-O5	C1-O1	C1-O5
	1,4	1.42101	1.4	1.41777	1.4	1.41351
	0.273135	0.261382	0.27197	0.259941	0.272293	0.265153
	C1	0.468135	C1	0.275119	C1	0.220731
	O5	-0.514638	O5	-0.335796	O5	-0.35199
	O1	-0.3874	O1	-0.369112	O1	-0.322083
1.45 Å	Bond Distance(Å)/Electron Density/Charge		Bond Distance(Å)/Electron Density/Charge		Bond Distance(Å)/Electron Density/Charge	
	C1-O1	C1-O5	C1-O1	C1-O5	C1-O1	C1-O5
	1.45	1.41204	1.45	1.40703	1.45	1.40457
	0.248914	0.266375	0.247641	0.265831	0.248046	0.270268
	C1	0.435398	C1	0.243241	C1	0.246717
	O5	-0.495095	O5	-0.317765	O5	-0.339593

	O1	-0.392516	O1	-0.375944	O1	-0.339958
1.50 Å	Bond Distance(Å)/Electron Density/Charge		Bond Distance(Å)/Electron Density/Charge		Bond Distance(Å)/Electron Density/Charge	
	C1-O1	C1-O5	C1-O1	C1-O5	C1-O1	C1-O5
	1.5	1.40413	1.5	1.39729	1.5	1.39645
	0.225997	0.270822	0.22466	0.271213	0.225157	0.274957
	C1	0.408451	C1	0.192364	C1	0.240308
	O5	-0.4869	O5	-0.302151	O5	-0.33027
	O1	-0.397351	O1	-0.376812	O1	-0.361211
1.55 Å	Bond Distance(Å)/Electron Density/Charge		Bond Distance(Å)/Electron Density/Charge		Bond Distance(Å)/Electron Density/Charge	
	C1-O1	C1-O5	C1-O1	C1-O5	C1-O1	C1-O5
	1.55	1.39704	1.55	1.38837	1.55	1.3891
	0.204422	0.274841	0.203026	0.276171	0.203622	0.27924
	C1	0.397478	C1	0.170484	C1	0.129829
	O5	-0.475039	O5	-0.287326	O5	-0.295274
	O1	-0.406692	O1	-0.387397	O1	-0.352259
1.60 Å	Bond Distance(Å)/Electron Density/Charge		Bond Distance(Å)/Electron Density/Charge		Bond Distance(Å)/Electron Density/Charge	
	C1-O1	C1-O5	C1-O1	C1-O5	C1-O1	C1-O5
	1.6	1.39057	1.6	1.38003	1.6	1.3825
	0.184382	0.278517	0.182928	0.280814	0.183622	0.283128
	C1	0.365046	C1	0.121897	C1	0.0895
	O5	-0.458803	O5	-0.263786	O5	-0.275009
	O1	-0.409624	O1	-0.402799	O1	-0.358178
1.65 Å	Bond Distance(Å)/Electron Density/Charge		Bond Distance(Å)/Electron Density/Charge		Bond Distance(Å)/Electron Density/Charge	
	C1-O1	C1-O5	C1-O1	C1-O5	C1-O1	C1-O5
	1.65	1.38462	1.65	1.37215	1.65	1.37643

	0.165976	0.281901	0.164465	0.285209	0.16525	0.286734
	C1	0.335741	C1	0.098191	C1	0.061005
	O5	-0.450869	O5	-0.248959	O5	-0.264864
	O1	-0.416085	O1	-0.41774	O1	-0.366349
1.70 Å	Bond Distance(Å)/Electron Density/Charge		Bond Distance(Å)/Electron Density/Charge		Bond Distance(Å)/Electron Density/Charge	
	C1-O1	C1-O5	C1-O1	C1-O5	C1-O1	C1-O5
	1.7	1.37907	1.7	1.36463	1.7	1.37071
	0.149211	0.285943	0.147647	0.289405	0.148516	0.290138
	C1	0.274728	C1	0.094398	C1	0.072298
	O5	-0.427704	O5	-0.241162	O5	-0.266265
	O1	-0.417362	O1	-0.427561	O1	-0.375904
1.75 Å	Bond Distance(Å)/Electron Density/Charge		Bond Distance(Å)/Electron Density/Charge		Bond Distance(Å)/Electron Density/Charge	
	C1-O1	C1-O5	C1-O1	C1-O5	C1-O1	C1-O5
	1.75	1.37384	1.75	1.35744	1.75	1.36531
	0.134045	0.287978	0.132431	0.293425	0.13338	0.293368
	C1	0.241095	C1	0.115334	C1	0.060255
	O5	-0.40598	O5	-0.279831	O5	-0.263716
	O1	-0.422428	O1	-0.441953	O1	-0.383567
1.80 Å	Bond Distance(Å)/Electron Density/Charge		Bond Distance(Å)/Electron Density/Charge		Bond Distance(Å)/Electron Density/Charge	
	C1-O1	C1-O5	C1-O1	C1-O5	C1-O1	C1-O5
	1.8	1.3688	1.8	1.35051	1.8	1.36021
	0.120401	0.290762	0.118751	0.297303	0.119759	0.296436
	C1	0.178681	C1	0.124667	C1	0.102784
	O5	-0.380229	O5	-0.260206	O5	-0.295756
	O1	-0.43264	O1	-0.459282	O1	-0.395705
1.85 Å	Bond Distance(Å)/Electron		Bond Distance(Å)/Electron		Bond Distance(Å)/Electron	

	Density/Charge		Density/Charge		Density/Charge	
	C1-O1	C1-O5	C1-O1	C1-O5 (oxo-carb)	C1-O1	C1-O5
	1.85	1.36387	1.85	1.34386	1.85	1.35537
	0.108177	0.293422	0.106511	0.301039	0.107574	0.299357
	C1	0.150936	C1	0.12482	C1	0.128651
	O5	-0.362934	O5	-0.253459	O5	-0.301103
	O1	-0.443692	O1	-0.476376	O1	-0.41096
1.90 Å	Bond Distance(Å)/Electron Density/Charge		Bond Distance(Å)/Electron Density/Charge		Bond Distance(Å)/Electron Density/Charge	
	C1-O1	C1-O5	C1-O1	C1-O5 (oxo-carb)	C1-O1	C1-O5
	1.9	1.35877	1.9	1.3375	1.9	1.35078
	0.0972515	0.296043	0.0956019	0.304629	0.0967088	0.302146
	C1	0.113866	C1	0.122893	C1	0.123093
	O5	-0.329785	O5	-0.247804	O5	-0.284305
	O1	-0.456493	O1	-0.488062	O1	-0.426602
2.00 Å	Bond Distance(Å)/Electron Density/Charge		Bond Distance(Å)/Electron Density/Charge		Bond Distance(Å)/Electron Density/Charge	
	C1-O1	C1-O5 (opt to 1S3-oc)	C1-O1	C1-O5 (oxo-carb)	C1-O1	C1-O5 (oxo-carb)
	2	1.32447	2	1.3258	2	1.34221
	0.0782452	0.308208	0.077294	0.311288	0.0748557	0.307394
	C1	0.235818	C1	0.108362	C1	0.135762
	O5	-0.392061	O5	-0.235416	O5	-0.289713
	O1	-0.606769	O1	-0.516162	O1	-0.450783
2.10 Å	Bond Distance(Å)/Electron Density/Charge		Bond Distance(Å)/Electron Density/Charge		Bond Distance(Å)/Electron Density/Charge	
	C1-O1	C1-O5 (opt to 1S3-oc)	C1-O1	C1-O5 (oxo-carb)	C1-O1	C1-O5 (oxo-carb)
	2.1	1.3163	2.1	1.31565	2.1	1.33441

	0.0639453	0.31276	0.0628705	0.317134	0.0640301	0.312202
	C1	0.251821	C1	0.101503	C1	0.174761
	O5	-0.386461	O5	-0.229962	O5	-0.312296
	O1	-0.631321	O1	-0.544518	O1	-0.472487
2.20 Å	Bond Distance(Å)/Electron Density/Charge		Bond Distance(Å)/Electron Density/Charge		Bond Distance(Å)/Electron Density/Charge	
	C1-O1	C1-O5 (opt to 4E-oc)	C1-O1	C1-O5 (oxo-carb)	C1-O1	C1-O5 (oxo-carb)
	2.2	1.30957	2.2	1.30721	2.2	1.3274
	C1	0.249162	C1	0.164305	C1	0.279718
	O5	-0.368381	O5	-0.241351	O5	-0.314557
	O1	-0.648775	O1	-0.586503	O1	-0.523247
2.30 Å	Bond Distance(Å)/Electron Density/Charge		Bond Distance(Å)/Electron Density/Charge		Bond Distance(Å)/Electron Density/Charge	
	C1-O1	C1-O5 (opt to 4E-oc)	C1-O1	C1-O5 (oxo-carb)	C1-O1	C1-O5 (oxo-carb)
	2.3	1.30413	2.3	1.30045	2.3	1.32115
	0.0423717	0.318664	0.0422806	0.326009	0.0433966	0.320226
	C1	0.244489	C1	0.143866	C1	0.354231
	O5	-0.349224	O5	-0.221844	O5	-0.33868
	O1	-0.671435	O1	-0.600899	O1	-0.555448
2.40 Å	Bond Distance(Å)/Electron Density/Charge		Bond Distance(Å)/Electron Density/Charge		Bond Distance(Å)/Electron Density/Charge	
	C1-O1	C1-O5 (opt to E3-oc)	C1-O1	C1-O5 (oxo-carb)	C1-O1	C1-O5 (oxo-carb)
	2.4	1.29742	2.4	1.29524	2.4	1.31562
	0.0345547	0.322777	0.0348614	0.329064	0.0359245	0.323473
	C1	0.34575	C1	0.154406	C1	0.426893
	O5	-0.371699	O5	-0.222702	O5	-0.347039
	O1	-0.615594	O1	-0.620596	O1	-0.589771

Table A.2 C1-O1, C1-O5 bond distances, electron densities and charges on C1, O5, O1 for $^{2}H_{3}$, ^{4}E, $^{4}H_{3}$

CONFORMER	$^{2}H_{3}$		^{4}E		$^{4}H_{3}$	
C1-O1 BOND LENGTH						
1.40 Å	Bond Distance(Å)/Electron Density/Charge		Bond Distance(Å)/Electron Density/Charge		Bond Distance(Å)/Electron Density/Charge	
	C1-O1	C1-O5	C1-O1	C1-O5	C1-O1	C1-O5
	1.4	1.42346	1.4	1.43377	1.4	1.42777
	0.272099	0.256579	0.273534	0.249206	0.273094	0.252201
	C1	0.163847	C1	0.303017	C1	0.432168
	O5	-0.33677	O5	-0.450437	O5	-0.507511
	O1	-0.36101	O1	-0.396837	O1	-0.445877
1.45 Å	Bond Distance(Å)/Electron Density/Charge		Bond Distance(Å)/Electron Density/Charge		Bond Distance(Å)/Electron Density/Charge	
	C1-O1	C1-O5	C1-O1	C1-O5	C1-O1	C1-O5
	1.45	1.41218	1.45	1.42105	1.45	1.41562
	0.247758	0.262672	0.249103	0.255814	0.248685	0.258484
	C1	0.124334	C1	0.262885	C1	0.410354
	O5	-0.335448	O5	-0.435863	O5	-0.493736
	O1	-0.362309	O1	-0.403502	O1	-0.451686
1.50 Å	Bond Distance(Å)/Electron Density/Charge		Bond Distance(Å)/Electron Density/Charge		Bond Distance(Å)/Electron Density/Charge	
	C1-O1	C1-O5	C1-O1	C1-O5	C1-O1	C1-O5
	1.5	1.40203	1.5	1.40959	1.5	1.40463
	0.224761	0.268203	0.225966	0.261817	0.225573	0.264191
	C1	0.099833	C1	0.232621	C1	0.367104
	O5	-0.314598	O5	-0.420433	O5	-0.477178

	O1	-0.376105	O1	-0.414909	O1	-0.458408
1.55 Å	Bond Distance(Å)/Electron Density/Charge		Bond Distance(Å)/Electron Density/Charge		Bond Distance(Å)/Electron Density/Charge	
	C1-O1	C1-O5	C1-O1	C1-O5	C1-O1	C1-O5
	1.55	1.39274	1.55	1.3992	1.55	1.39447
	0.203115	0.273286	0.204186	0.267296	0.203822	0.269471
	C1	0.077672	C1	0.181333	C1	0.338872
	O5	-0.30258	O5	-0.395707	O5	-0.467404
	O1	-0.385852	O1	-0.427446	O1	-0.466142
1.60 Å	Bond Distance(Å)/Electron Density/Charge		Bond Distance(Å)/Electron Density/Charge		Bond Distance(Å)/Electron Density/Charge	
	C1-O1	C1-O5	C1-O1	C1-O5	C1-O1	C1-O5
	1.6	1.38409	1.6	1.38954	1.6	1.38525
	0.183008	0.278035	0.183965	0.272408	0.183626	0.274275
	C1	0.095379	C1	0.172902	C1	0.320233
	O5	-0.30603	O5	-0.385505	O5	-0.453747
	O1	-0.406681	O1	-0.436879	O1	-0.481272
1.65 Å	Bond Distance(Å)/Electron Density/Charge		Bond Distance(Å)/Electron Density/Charge		Bond Distance(Å)/Electron Density/Charge	
	C1-O1	C1-O5	C1-O1	C1-O5	C1-O1	C1-O5
	1.65	1.37592	1.65	1.38055	1.65	1.37654
	0.164538	0.282523	0.165398	0.277187	0.165091	0.27882
	C1	0.080367	C1	0.173262	C1	0.313766
	O5	-0.296801	O5	-0.36951	O5	-0.44609
	O1	-0.419359	O1	-0.45627	O1	-0.498037
1.70 Å	Bond Distance(Å)/Electron Density/Charge		Bond Distance(Å)/Electron Density/Charge		Bond Distance(Å)/Electron Density/Charge	
	C1-O1	C1-O5	C1-O1	C1-O5	C1-O1	C1-O5
	1.7	1.36814	1.7	1.3722	1.7	1.36838

	0.147715	0.286799	0.148499	0.28164	0.148223	0.283081
	C1	0.074351	C1	0.204466	C1	0.293304
	O5	-0.291233	O5	-0.371624	O5	-0.439711
	O1	-0.433645	O1	-0.471313	O1	-0.507954
1.75 Å	Bond Distance(Å)/Electron Density/Charge		Bond Distance(Å)/Electron Density/Charge		Bond Distance(Å)/Electron Density/Charge	
	C1-O1	C1-O5	C1-O1	C1-O5	C1-O1	C1-O5
	1.75	1.36072	1.75	1.36412	1.75	1.36079
	0.132501	0.290891	0.133188	0.285991	0.132977	0.287061
	C1	0.05513	C1	0.359795	C1	0.269322
	O5	-0.283275	O5	-0.409013	O5	-0.427981
	O1	-0.444066	O1	-0.488765	O1	-0.518609
1.80 Å	Bond Distance(Å)/Electron Density/Charge		Bond Distance(Å)/Electron Density/Charge		Bond Distance(Å)/Electron Density/Charge	
	C1-O1	C1-O5	C1-O1	C1-O5	C1-O1	C1-O5
	1.8	1.35366	1.8	1.35652	1.8	1.35371
	0.118813	0.294789	0.11945	0.290134	0.119286	0.290788
	C1	0.046038	C1	0.324801	C1	0.257395
	O5	-0.278094	O5	-0.395823	O5	-0.414813
	O1	-0.457826	O1	-0.496475	O1	-0.531144
1.85 Å	Bond Distance(Å)/Electron Density/Charge		Bond Distance(Å)/Electron Density/Charge		Bond Distance(Å)/Electron Density/Charge	
	C1-O1	C1-O5	C1-O1	C1-O5	C1-O1	C1-O5
	1.85	1.34693	1.85	1.34966	1.85	1.34717
	0.106572	0.298526	0.107158	0.293409	0.10705	0.294254
	C1	0.033316	C1	0.220273	C1	0.2462
	O5	-0.269718	O5	-0.352386	O5	-0.397978
	O1	-0.47345	O1	-0.509332	O1	-0.545327
1.90 Å	Bond Distance(Å)/Electron		Bond Distance(Å)/Electron		Bond Distance(Å)/Electron	

	Density/Charge		Density/Charge		Density/Charge	
	C1-O1	C1-O5 (oxo-carb)	C1-O1	C1-O5 (oxo-carb)	C1-O1	C1-O5
	1.9	1.34054	1.9	1.34313	1.9	1.34116
	0.0956664	0.302085	0.0961873	0.297482	0.0961574	0.297458
	C1	0.048264	C1	0.365148	C1	0.24712
	O5	-0.268357	O5	-0.384746	O5	-0.393333
	O1	-0.49121	O1	-0.538937	O1	-0.568614
2.00 Å	Bond Distance(Å)/Electron Density/Charge		Bond Distance(Å)/Electron Density/Charge		Bond Distance(Å)/Electron Density/Charge	
	C1-O1	C1-O5 (oxo-carb)	C1-O1	C1-O5 (oxo-carb)	C1-O1	C1-O5 (oxo-carb)
	2	1.32892	2	1.33118	2	1.33064
	0.073723	0.308604	0.077417	0.304202	0.0790836	0.303136
	C1	0.044957	C1	0.417808	C1	0.150991
	O5	-0.264832	O5	-0.385046	O5	-0.362935
	O1	-0.521455	O1	-0.586601	O1	-0.581039
2.10 Å	Bond Distance(Å)/Electron Density/Charge		Bond Distance(Å)/Electron Density/Charge		Bond Distance(Å)/Electron Density/Charge	
	C1-O1	C1-O5 (oxo-carb)	C1-O1	C1-O5 (oxo-carb)	C1-O1	C1-O5 (oxo-carb)
	2.1	1.31902	2.1	1.32092	2.1	1.32181
	0.0629621	0.314213	0.0632114	0.310208	0.0635183	0.308037
	C1	0.072057	C1	0.430499	C1	0.201422
	O5	-0.257344	O5	-0.364152	O5	-0.375099
	O1	-0.55244	O1	-0.620604	O1	-0.612598
2.20 Å	Bond Distance(Å)/Electron Density/Charge		Bond Distance(Å)/Electron Density/Charge		Bond Distance(Å)/Electron Density/Charge	
	C1-O1	C1-O5	C1-O1	C1-O5	C1-O1	C1-O5

		(oxo-carb)		(oxo-carb)		(oxo-carb)
	2.2	1.31097	2.2	1.31242	2.2	1.31434
	0.0515233	0.318802	0.051759	0.315347	0.0520822	0.312359
	C1	0.0959	C1	0.446655	C1	0.215232
	O5	-0.264771	O5	-0.356998	O5	-0.357279
	O1	-0.585183	O1	-0.642185	O1	-0.63256
2.30 Å	Bond Distance(Å)/Electron Density/Charge		Bond Distance(Å)/Electron Density/Charge		Bond Distance(Å)/Electron Density/Charge	
	C1-O1	C1-O5 (oxo-carb)	C1-O1	C1-O5 (oxo-carb)	C1-O1	C1-O5
	2.3	1.30444	2.3	1.30547	2.3	1.3082
	0.0423598	0.322543	0.0426625	0.319694	0.0428978	0.316017
	C1	0.121243	C1	0.42007	C1	0.244756
	O5	-0.256714	O5	-0.339668	O5	-0.382916
	O1	-0.602866	O1	-0.662452	O1	-0.654383
2.40 Å	Bond Distance(Å)/Electron Density/Charge		Bond Distance(Å)/Electron Density/Charge		Bond Distance(Å)/Electron Density/Charge	
	C1-O1	C1-O5 (oxo-carb)	C1-O1	C1-O5 (oxo-carb)	C1-O1	C1-O5
	2.4	1.29892	2.4	1.29982	2.4	1.30332
	0.0349118	0.325752	0.0353641	0.323377	0.0353817	0.318965
	C1	0.206824	C1	0.361985	C1	0.273287
	O5	-0.267831	O5	-0.305766	O5	-0.378497
	O1	-0.634053	O1	-0.693784	O1	-0.668625

Table A.3 C1-O1, C1-O5 bond distances, electron densities and charges on C1, O5, O1 for $^4H_5$, E_1, E_3

CONFORMER	$^4H_5$		E_1		E_3	
C1-O1 BOND LENGTH						
1.40 Å	Bond Distance(Å)/Electron Density/Charge		Bond Distance(Å)/Electron Density/Charge		Bond Distance(Å)/Electron Density/Charge	
	C1-O1	C1-O5	C1-O1	C1-O5	C1-O1	C1-O5
	1.4	1.4361	1.4	1.42657	1.4	1.42553
	0.274197	0.250712	0.272105	0.260121	0.272034	0.253263
	C1	0.356716	C1	0.349933	C1	0.309914
	O5	-0.416506	O5	-0.407948	O5	-0.451192
	O1	-0.408696	O1	-0.341888	O1	-0.387289
1.45 Å	Bond Distance(Å)/Electron Density/Charge		Bond Distance(Å)/Electron Density/Charge		Bond Distance(Å)/Electron Density/Charge	
	C1-O1	C1-O5	C1-O1	C1-O5	C1-O1	C1-O5
	1.45	1.42507	1.45	1.41646	1.45	1.41309
	0.249833	0.256535	0.247887	0.265736	0.247651	0.259793
	C1	0.322593	C1	0.314588	C1	0.281814
	O5	-0.399792	O5	-0.389786	O5	-0.425989
	O1	-0.407346	O1	-0.345383	O1	-0.401899
1.50 Å	Bond Distance(Å)/Electron Density/Charge		Bond Distance(Å)/Electron Density/Charge		Bond Distance(Å)/Electron Density/Charge	
	C1-O1	C1-O5	C1-O1	C1-O5	C1-O1	C1-O5
	1.5	1.41517	1.5	1.40759	1.5	1.40178
	0.226742	0.2618001	0.225013	0.270732	0.224597	0.265766
	C1	0.212977	C1	0.313493	C1	0.21784
	O5	-0.384859	O5	-0.390934	O5	-0.40809

	O1	-0.393763	O1	-0.352172	O1	-0.397352
1.55 Å	Bond Distance(Å)/Electron Density/Charge		Bond Distance(Å)/Electron Density/Charge		Bond Distance(Å)/Electron Density/Charge	
	C1-O1	C1-O5	C1-O1	C1-O5	C1-O1	C1-O5
	1.55	1.40614	1.55	1.39978	1.55	1.39129
	0.204992	0.266623	0.203505	0.275185	0.202899	0.271337
	C1	0.194816	C1	0.292723	C1	0.148737
	O5	-0.375481	O5	-0.379476	O5	-0.383628
	O1	-0.402825	O1	-0.353855	O1	-0.398846
1.60 Å	Bond Distance(Å)/Electron Density/Charge		Bond Distance(Å)/Electron Density/Charge		Bond Distance(Å)/Electron Density/Charge	
	C1-O1	C1-O5	C1-O1	C1-O5	C1-O1	C1-O5
	1.6	1.39774	1.6	1.3927	1.6	1.38167
	0.184789	0.271125	0.183546	0.279239	0.182749	0.276478
	C1	0.149083	C1	0.245912	C1	0.151492
	O5	-0.349458	O5	-0.364709	O5	-0.364578
	O1	-0.406288	O1	-0.350286	O1	-0.416118
1.65 Å	Bond Distance(Å)/Electron Density/Charge		Bond Distance(Å)/Electron Density/Charge		Bond Distance(Å)/Electron Density/Charge	
	C1-O1	C1-O5	C1-O1	C1-O5	C1-O1	C1-O5
	1.65	1.38981	1.65	1.3862	1.65	1.37253
	0.166227	0.275373	0.165211	0.282988	0.164251	0.281391
	C1	0.167677	C1	0.244126	C1	0.14361
	O5	-0.349419	O5	-0.347034	O5	-0.349961
	O1	-0.430382	O1	-0.359437	O1	-0.432852
1.70 Å	Bond Distance(Å)/Electron Density/Charge		Bond Distance(Å)/Electron Density/Charge		Bond Distance(Å)/Electron Density/Charge	
	C1-O1	C1-O5	C1-O1	C1-O5	C1-O1	C1-O5
	1.7	1.38226	1.7	1.38026	1.7	1.36392

	0.149317	0.279415	0.148513	0.286429	0.147411	0.286049
	C1	0.157233	C1	0.311105	C1	0.141554
	O5	-0.342696	O5	-0.352311	O5	-0.358294
	O1	-0.444451	O1	-0.398019	O1	-0.448538
1.75 Å	Bond Distance(Å)/Electron Density/Charge		Bond Distance(Å)/Electron Density/Charge		Bond Distance(Å)/Electron Density/Charge	
	C1-O1	C1-O5	C1-O1	C1-O5	C1-O1	C1-O5
	1.75	1.37507	1.75	1.37464	1.75	1.3558
	0.134007	0.283271	0.133409	0.289708	0.132193	0.290487
	C1	0.162154	C1	0.252819	C1	0.163443
	O5	-0.340668	O5	-0.333168	O5	-0.343182
	O1	-0.458743	O1	-0.394197	O1	-0.466219
1.80 Å	Bond Distance(Å)/Electron Density/Charge		Bond Distance(Å)/Electron Density/Charge		Bond Distance(Å)/Electron Density/Charge	
	C1-O1	C1-O5	C1-O1	C1-O5	C1-O1	C1-O5
	1.8	1.3682	1.8	1.36931	1.8	1.34815
	0.120238	0.28695	0.119834	0.292813	0.118521	0.294715
	C1	0.189333	C1	0.207065	C1	0.161052
	O5	-0.349097	O5	-0.303703	O5	-0.337278
	O1	-0.471933	O1	-0.403358	O1	-0.48593
1.85 Å	Bond Distance(Å)/Electron Density/Charge		Bond Distance(Å)/Electron Density/Charge		Bond Distance(Å)/Electron Density/Charge	
	C1-O1	C1-O5	C1-O1	C1-O5	C1-O1	C1-O5 (oxo-carb)
	1.85	1.36166	1.85	1.36437	1.85	1.34094
	0.107908	0.290439	0.107691	0.295698	0.106296	0.298766
	C1	0.250192	C1	0.21728	C1	0.160524
	O5	-0.350801	O5	-0.303605	O5	-0.333378
	O1	-0.495485	O1	-0.414478	O1	-0.498017
1.90 Å	Bond		Bond		Bond	

	Distance(Å)/Electron Density/Charge		Distance(Å)/Electron Density/Charge		Distance(Å)/Electron Density/Charge	
	C1-O1	C1-O5	C1-O1	C1-O5	C1-O1	C1-O5
	1.9	1.35547	1.9	1.35986	1.9	1.33411
	0.0969092	0.293735	0.0968701	0.298338	0.0954024	0.302697
	C1	0.254237	C1	0.219539	C1	0.167729
	O5	-0.340961	O5	-0.30845	O5	-0.325072
	O1	-0.504488	O1	-0.419607	O1	-0.505813
2.00 Å	Bond Distance(Å)/Electron Density/Charge		Bond Distance(Å)/Electron Density/Charge		Bond Distance(Å)/Electron Density/Charge	
	C1-O1	C1-O5 (oxo-carb)	C1-O1	C1-O5	C1-O1	C1-O5
	2	1.3438	2	1.35184	2	1.32128
	0.0784112	0.299999	0.0787184	0.303111	0.0770593	0.310436
	C1	0.28632	C1	0.207331	C1	0.142087
	O5	-0.33309	O5	-0.294791	O5	-0.27779
	O1	-0.525486	O1	-0.429772	O1	-0.526586
2.10 Å	Bond Distance(Å)/Electron Density/Charge		Bond Distance(Å)/Electron Density/Charge		Bond Distance(Å)/Electron Density/Charge	
	C1-O1	C1-O5	C1-O1	C1-O5	C1-O1	C1-O5
	2.1	1.33326	2.1	1.34398	2.1	1.30998
	0.0637804	0.305688	0.0643307	0.308349	0.0625439	0.31755
	C1	0.362458	C1	0.252235	C1	0.156302
	O5	-0.344847	O5	-0.33062	O5	-0.267067
	O1	-0.596381	O1	-0.477312	O1	-0.558257
2.20 Å	Bond Distance(Å)/Electron Density/Charge		Bond Distance(Å)/Electron Density/Charge		Bond Distance(Å)/Electron Density/Charge	
	C1-O1	C1-O5	C1-O1	C1-O5	C1-O1	C1-O5

	2.2	1.32433	2.2	1.33787	2.2	1.30061
	0.0521993	0.310535	0.052959	0.312211	0.0510383	0.323649
	C1	0.418195	C1	0.32277	C1	0.138876
	O5	-0.34002	O5	-0.360482	O5	-0.241872
	O1	-0.6393	O1	-0.514338	O1	-0.581147
2.30 Å	Bond Distance(Å)/Electron Density/Charge		Bond Distance(Å)/Electron Density/Charge		Bond Distance(Å)/Electron Density/Charge	
	C1-O1	C1-O5	C1-O1	C1-O5	C1-O1	C1-O5
	2.3	1.31652	2.3	1.33277	2.3	1.29314
	0.0429523	0.314838	0.0438574	0.315404	0.0418357	0.328597
	C1	0.400061	C1	0.335665	C1	0.156458
	O5	-0.318491	O5	-0.364376	O5	-0.239979
	O1	-0.694843	O1	-0.543135	O1	-0.606902
2.40 Å	Bond Distance(Å)/Electron Density/Charge		Bond Distance(Å)/Electron Density/Charge		Bond Distance(Å)/Electron Density/Charge	
	C1-O1	C1-O5	C1-O1	C1-O5	C1-O1	C1-O5
	2.4	1.31065	2.4	1.32853	2.4	1.28701
	0.035537	0.318129	0.0364691	0.318052	0.0343556	0.332725
	C1	0.412822	C1	0.396136	C1	0.143045
	O5	-0.313302	O5	-0.385782	O5	-0.231776
	O1	-0.736619	O1	-0.58534	O1	-0.613262

Table A.4 C1-O1, C1-O5 bond distances, electron densities and charges on C1, O5, O1 for E_5, ^{O}E, $^{O}H_1$

CONFORMER	E_5		^{O}E		$^{O}H_1$	
C1-O1 BOND LENGTH						
1.40 Å	Bond Distance(Å)/Electron Density/Charge		Bond Distance(Å)/Electron Density/Charge		Bond Distance(Å)/Electron Density/Charge	
	C1-O1	C1-O5	C1-O1	C1-O5	C1-O1	C1-O5
	1.4	1.42744	1.4	1.4214	1.4	1.43801
	0.274414	0.252742	0.272885	0.259943	0.272058	0.254377
	C1	0.535674	C1	0.5422	C1	0.261644
	O5	-0.560418	O5	-0.524084	O5	-0.411635
	O1	-0.449888	O1	-0.431604	O1	-0.332872
1.45 Å	Bond Distance(Å)/Electron Density/Charge		Bond Distance(Å)/Electron Density/Charge		Bond Distance(Å)/Electron Density/Charge	
	C1-O1	C1-O5	C1-O1	C1-O5	C1-O1	C1-O5
	1.45	1.4165	1.45	1.41261	1.45	1.42762
	0.250004	0.258544	0.248756	0.264837	0.247971	0.259999
	C1	0.545882	C1	0.506072	C1	0.301141
	O5	-0.546108	O5	-0.50882	O5	-0.401091
	O1	-0.465089	O1	-0.431289	O1	-0.353222
1.50 Å	Bond Distance(Å)/Electron Density/Charge		Bond Distance(Å)/Electron Density/Charge		Bond Distance(Å)/Electron Density/Charge	
	C1-O1	C1-O5	C1-O1	C1-O5	C1-O1	C1-O5
	1.5	1.40669	1.5	1.40486	1.5	1.41867
	0.226866	0.263787	0.225935	0.269206	0.225166	0.264899
	C1	0.550127	C1	0.485473	C1	0.225132
	O5	-0.54683	O5	-0.50047	O5	-0.381105

	O1	-0.47495	O1	-0.438163	O1	-0.355592
1.55 Å	Bond Distance(Å)/Electron Density/Charge		Bond Distance(Å)/Electron Density/Charge		Bond Distance(Å)/Electron Density/Charge	
	C1-O1	C1-O5	C1-O1	C1-O5	C1-O1	C1-O5
	1.55	1.3978	1.55	1.39793	1.55	1.41063
	0.205075	0.268572	0.204443	0.273146	0.203708	0.269333
	C1	0.551765	C1	0.457166	C1	0.194457
	O5	-0.550341	O5	-0.492318	O5	-0.384737
	O1	-0.483281	O1	-0.441865	O1	-0.363278
1.60 Å	Bond Distance(Å)/Electron Density/Charge		Bond Distance(Å)/Electron Density/Charge		Bond Distance(Å)/Electron Density/Charge	
	C1-O1	C1-O5	C1-O1	C1-O5	C1-O1	C1-O5
	1.6	1.38952	1.6	1.39181	1.6	1.4034
	0.184837	0.273032	0.184469	0.276666	0.183773	0.273329
	C1	0.54771	C1	0.41768	C1	0.165498
	O5	-0.558058	O5	-0.483683	O5	-0.375907
	O1	-0.48889	O1	-0.446247	O1	-0.368636
1.65 Å	Bond Distance(Å)/Electron Density/Charge		Bond Distance(Å)/Electron Density/Charge		Bond Distance(Å)/Electron Density/Charge	
	C1-O1	C1-O5	C1-O1	C1-O5	C1-O1	C1-O5
	1.65	1.38174	1.65	1.38619	1.65	1.39687
	0.16625	0.277222	0.166119	0.279912	0.165457	0.276969
	C1	0.476935	C1	0.41252	C1	0.118271
	O5	-0.522673	O5	-0.45933	O5	-0.351348
	O1	-0.489498	O1	-0.461985	O1	-0.37512
1.70 Å	Bond Distance(Å)/Electron Density/Charge		Bond Distance(Å)/Electron Density/Charge		Bond Distance(Å)/Electron Density/Charge	
	C1-O1	C1-O5	C1-O1	C1-O5	C1-O1	C1-O5
	1.7	1.37438	1.7	1.38103	1.7	1.39073

	0.149322	0.281184	0.149401	0.282901	0.148771	0.280391
	C1	0.479863	C1	0.389205	C1	0.152967
	O5	-0.51769	O5	-0.450109	O5	-0.34532
	O1	-0.503776	O1	-0.468715	O1	-0.382142
1.75 Å	Bond Distance(Å)/Electron Density/Charge		Bond Distance(Å)/Electron Density/Charge		Bond Distance(Å)/Electron Density/Charge	
	C1-O1	C1-O5	C1-O1	C1-O5	C1-O1	C1-O5
	1.75	1.36739	1.75	1.37626	1.75	1.38516
	0.134001	0.284948	0.134275	0.285671	0.133643	0.283518
	C1	0.449974	C1	0.362167	C1	0.202071
	O5	-0.488922	O5	-0.436351	O5	-0.34977
	O1	-0.510346	O1	-0.47254	O1	-0.401186
1.80 Å	Bond Distance(Å)/Electron Density/Charge		Bond Distance(Å)/Electron Density/Charge		Bond Distance(Å)/Electron Density/Charge	
	C1-O1	C1-O5	C1-O1	C1-O5	C1-O1	C1-O5
	1.8	1.36074	1.8	1.37183	1.8	1.37997
	0.120227	0.288526	0.120661	0.288254	0.120047	0.286434
	C1	0.478515	C1	0.328846	C1	0.182724
	O5	-0.492213	O5	-0.422901	O5	-0.332076
	O1	-0.517649	O1	-0.478095	O1	-0.410512
1.85 Å	Bond Distance(Å)/Electron Density/Charge		Bond Distance(Å)/Electron Density/Charge		Bond Distance(Å)/Electron Density/Charge	
	C1-O1	C1-O5	C1-O1	C1-O5	C1-O1	C1-O5
	1.85	1.35441	1.85	1.36768	1.85	1.37516
	0.107895	0.291921	0.108477	0.290667	0.107891	0.289132
	C1	0.471242	C1	0.312834	C1	0.158291
	O5	-0.480902	O5	-0.417781	O5	-0.322572
	O1	-0.525109	O1	-0.48422	O1	-0.415432
1.90 Å	Bond Distance(Å)/Electron		Bond Distance(Å)/Electron		Bond Distance(Å)/Electron	

	Density/Charge		Density/Charge		Density/Charge	
	C1-O1	C1-O5	C1-O1	C1-O5	C1-O1	C1-O5
	1.9	1.34843	1.9	1.36379	1.9	1.37074
	0.0968945	0.295131	0.0976099	0.292928	0.0970604	0.291624
	C1	0.475368	C1	0.310732	C1	0.138291
	O5	-0.477573	O5	-0.417339	O5	-0.308895
	O1	-0.538816	O1	-0.490495	O1	-0.425634
2.00 Å	Bond Distance(Å)/Electron Density/Charge		Bond Distance(Å)/Electron Density/Charge		Bond Distance(Å)/Electron Density/Charge	
	C1-O1	C1-O5	C1-O1	C1-O5	C1-O1	C1-O5
	2	1.33758	2	1.35671	2	1.36287
	0.0784106	0.300991	0.0793483	0.297038	0.0788585	0.296131
	C1	0.513774	C1	0.256492	C1	0.109578
	O5	-0.48327	O5	-0.387239	O5	-0.293193
	O1	-0.576264	O1	-0.505879	O1	-0.439295
2.10 Å	Bond Distance(Å)/Electron Density/Charge		Bond Distance(Å)/Electron Density/Charge		Bond Distance(Å)/Electron Density/Charge	
	C1-O1	C1-O5	C1-O1	C1-O5	C1-O1	C1-O5
	2.1	1.32825	2.1	1.35047	2.1	1.35615
	0.063841	0.305966	0.0649323	0.300607	0.0644371	0.300199
	C1	0.545772	C1	0.288438	C1	0.092187
	O5	-0.475882	O5	-0.366899	O5	-0.295084
	O1	-0.628337	O1	-0.530967	O1	-0.459554
2.20 Å	Bond Distance(Å)/Electron Density/Charge		Bond Distance(Å)/Electron Density/Charge		Bond Distance(Å)/Electron Density/Charge	
	C1-O1	C1-O5	C1-O1	C1-O5	C1-O1	C1-O5
	2.2	1.32032	2.2	1.34503	2.2	1.35047

	0.052323	0.310256	0.053493	0.303675	0.0530005	0.30365
	C1	0.557188	C1	0.291924	C1	0.177194
	O5	-0.460136	O5	-0.359733	O5	-0.335523
	O1	-0.663753	O1	-0.560741	O1	-0.48498
2.30 Å	Bond Distance(Å)/Electron Density/Charge		Bond Distance(Å)/Electron Density/Charge		Bond Distance(Å)/Electron Density/Charge	
	C1-O1	C1-O5	C1-O1	C1-O5	C1-O1	C1-O5
	2.3	1.31368	2.3	1.3402	2.3	1.34556
	0.0430975	0.313733	0.0442959	0.306409	0.0437815	0.306674
	C1	0.541379	C1	0.390175	C1	0.192263
	O5	-0.404348	O5	-0.376519	O5	-0.343819
	O1	-0.695311	O1	-0.596371	O1	-0.521387
2.40 Å	Bond Distance(Å)/Electron Density/Charge		Bond Distance(Å)/Electron Density/Charge		Bond Distance(Å)/Electron Density/Charge	
	C1-O1	C1-O5	C1-O1	C1-O5	C1-O1	C1-O5
	2.4	1.30846	2.4	1.33587	2.4	1.34131
	0.0356773	0.316208	0.0368046	0.308858	0.0362769	0.309317
	C1	0.539222	C1	0.419293	C1	0.297849
	O5	-0.384093	O5	-0.394558	O5	-0.372728
	O1	-0.724639	O1	-0.603252	O1	-0.562605

Table A.5 C1-O1, C1-O5 bond distances, electron densities and charges on C1, O5, O1 for $^{O}H_5$

CONFORMER	$^{O}H_5$	
C1-O1 BOND LENGTH		
1.40 Å	Bond Distance(Å)/Electron Density/Charge	
	C1-O1	C1-O5
	1.4	1.4182
	0.272955	0.263281
	C1	0.403586
	O5	-0.446234
	O1	-0.384261
1.45 Å	Bond Distance(Å)/Electron Density/Charge	
	C1-O1	C1-O5
	1.45	1.40959
	0.248832	0.268144
	C1	0.367219
	O5	-0.425612
	O1	-0.390309
1.50 Å	Bond Distance(Å)/Electron Density/Charge	
	C1-O1	C1-O5
	1.5	1.40197
	0.22602	0.272481
	C1	0.338406
	O5	-0.408674
	O1	-0.396955
1.55 Å	Bond Distance(Å)/Electron	

	Density/Charge	
	C1-O1	C1-O5
	1.55	1.39518
	0.204531	0.276381
	C1	0.28591
	O5	-0.390677
	O1	-0.395215
1.60 Å	Bond Distance(Å)/Electron Density/Charge	
	C1-O1	C1-O5
	1.6	1.38906
	0.184561	0.279926
	C1	0.2631
	O5	-0.379735
	O1	-0.402046
1.65 Å	Bond Distance(Å)/Electron Density/Charge	
	C1-O1	C1-O5
	1.65	1.38351
	0.166207	0.28317
	C1	0.237899
	O5	-0.360953
	O1	-0.410489
1.70 Å	Bond Distance(Å)/Electron Density/Charge	
	C1-O1	C1-O5
	1.7	1.37843
	0.149482	0.286151
	C1	0.167571
	O5	-0.340451

	O1	-0.405715
1.75 Å	Bond Distance(Å)/Electron Density/Charge	
	C1-O1	C1-O5
	1.75	1.3738
	0.134341	0.288888
	C1	0.132914
	O5	-0.323295
	O1	-0.41129
1.80 Å	Bond Distance(Å)/Electron Density/Charge	
	C1-O1	C1-O5
	1.8	1.36953
	0.120727	0.291411
	C1	0.11487
	O5	-0.319275
	O1	-0.421362
1.85 Å	Bond Distance(Å)/Electron Density/Charge	
	C1-O1	C1-O5
	1.85	1.36555
	0.108544	0.293757
	C1	0.079576
	O5	-0.310368
	O1	-0.42305
1.90 Å	Bond Distance(Å)/Electron Density/Charge	
	C1-O1	C1-O5
	1.9	1.36182
	0.0976801	0.295957
	C1	0.052069

	O5	-0.29584
	O1	-0.433905
2.00 Å	Bond Distance(Å)/Electron Density/Charge	
	C1-O1	C1-O5
	2	1.35497
	0.079424	0.3001
	C1	0.089019
	O5	-0.295358
	O1	-0.474211
2.10 Å	Bond Distance(Å)/Electron Density/Charge	
	C1-O1	C1-O5
	2.1	1.34893
	0.0650041	0.303608
	C1	0.155687
	O5	-0.268064
	O1	-0.507151
2.20 Å	Bond Distance(Å)/Electron Density/Charge	
	C1-O1	C1-O5
	2.2	1.34372
	0.0535485	0.306699
	C1	0.179186
	O5	-0.254115
	O1	-0.534941
2.30 Å	Bond Distance(Å)/Electron Density/Charge	
	C1-O1	C1-O5
	2.3	1.3391
	0.0443302	0.309477

	C1	0.150914
	O5	-0.236594
	O1	-0.549887
2.40 Å	Bond Distance(Å)/Electron Density/Charge	
	C1-O1	C1-O5
	2.4	1.33506
	0.368308	0.311901
	C1	0.161155
	O5	-0.252864
	O1	-0.566613

APPENDIX B

Figures of β-D-Methyl Glucose for conformers $^{4}E_{5}$, $^{4}H_{5}$, ^{4}E, $^{4}H_{3}$, E_{3}, $^{2}H_{3}$, ^{2}E, $^{2}H_{1}$, E_{1}, $^{O}H_{1}$, ^{O}E and $^{O}H_{5}$

B.1 Electron density across the C1-O5 bond, charge on C1, O1 and O5

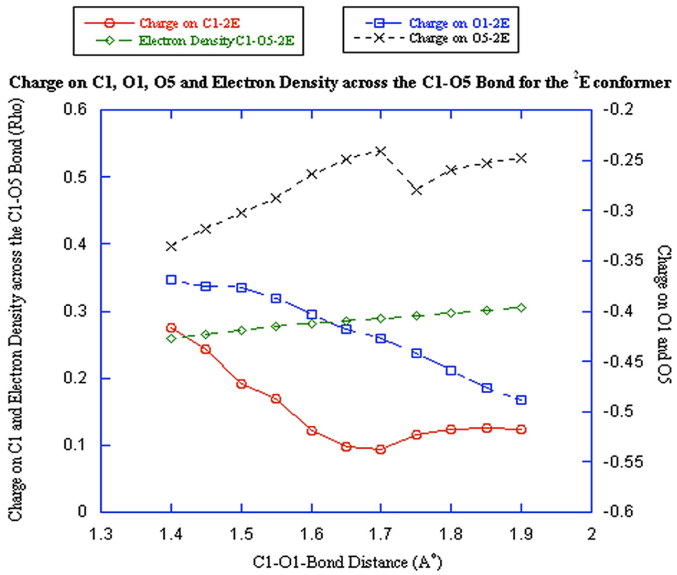


Figure B.1 The electron density across the C1-O5 bond, charge on C1, O1 and O5 as a function of the C1-OMe bond length for the ^{2}E conformer

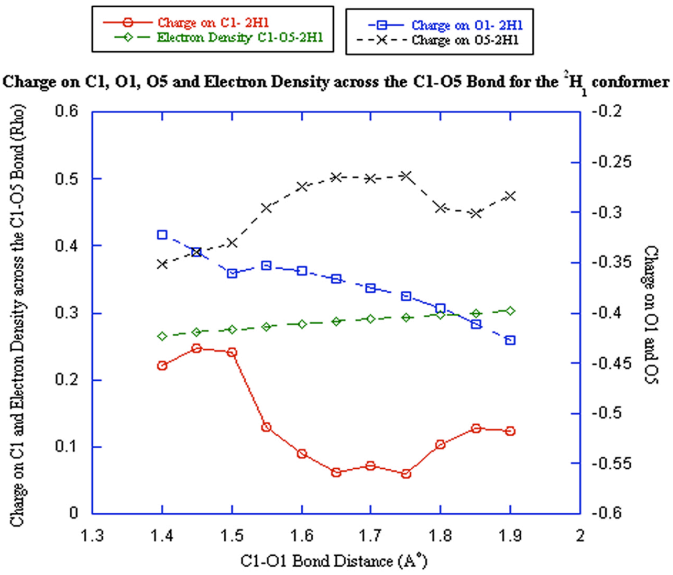


Figure B.2 The electron density across the C1-O5 bond, charge on C1, O1 and O5 as a function of the C1-OMe bond length for the $^2H_1$ conformer

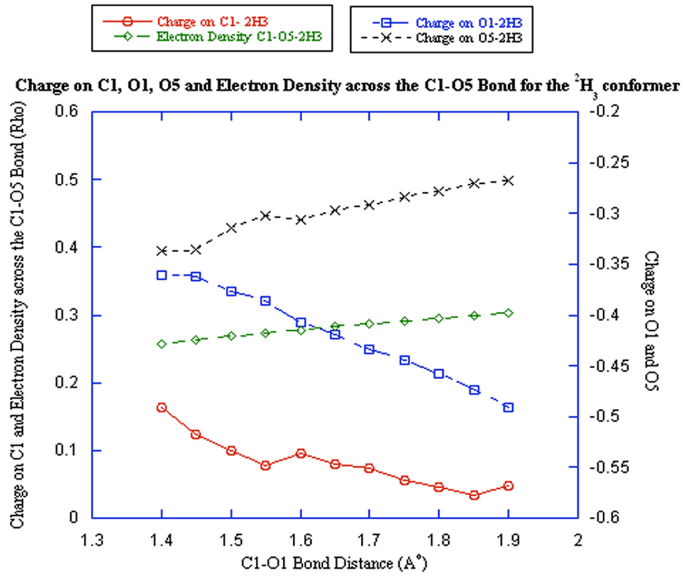


Figure B.3 The electron density across the C1-O5 bond, charge on C1, O1 and O5 as a function of the C1-OMe bond length for the $^2H_3$ conformer

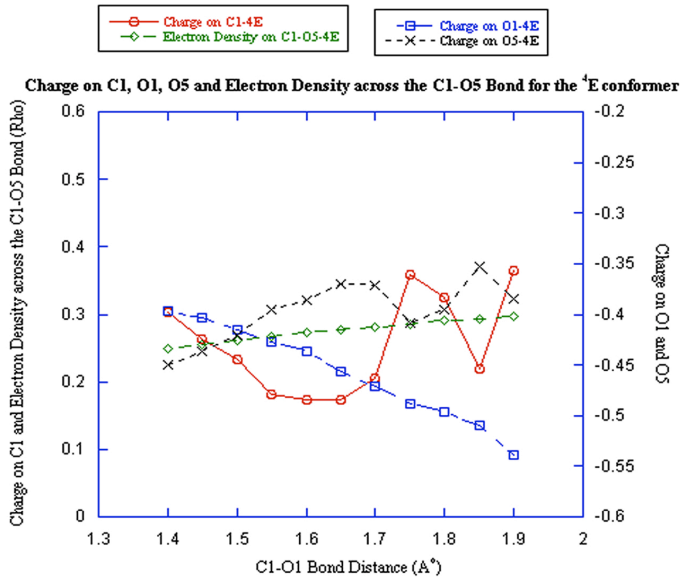


Figure B.4 The electron density across the C1-O5 bond, charge on C1, O1 and O5 as a function of the C1-OMe bond length for the ^{4}E conformer

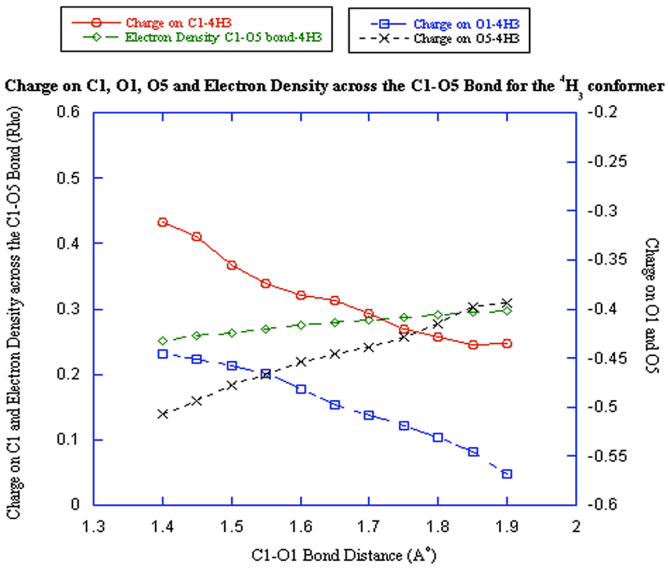


Figure B.5 The electron density across the C1-O5 bond, charge on C1, O1 and O5 as a function of the C1-OMe bond length for the $^{4}H_{3}$ conformer

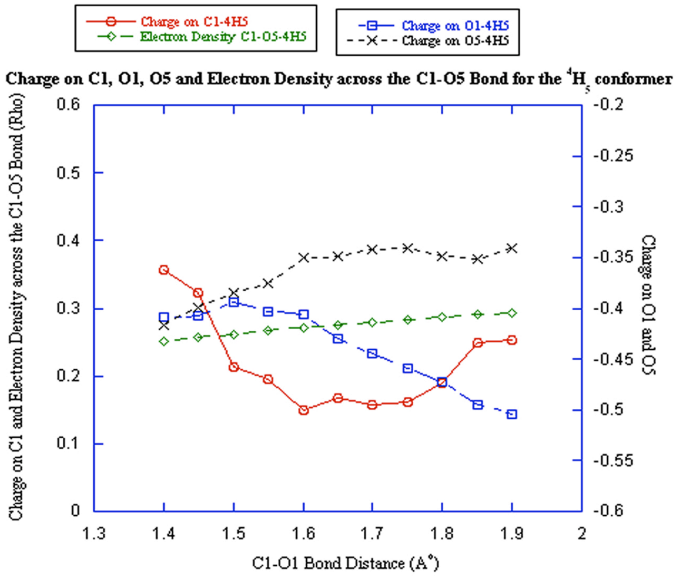


Figure B.6 The electron density across the C1-O5 bond, charge on C1, O1 and O5 as a function of the C1-OMe bond length for the $^4H_5$ conformer

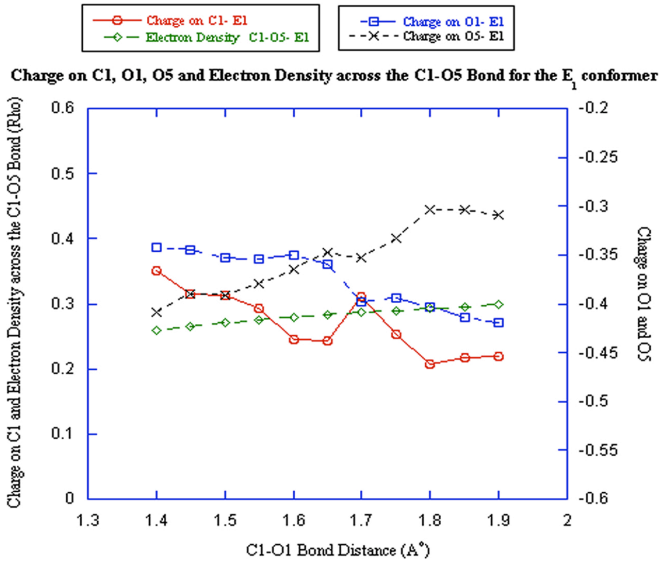


Figure B.7 The electron density across the C1-O5 bond, charge on C1, O1 and O5 as a function of the C1-OMe bond length for the E_1 conformer

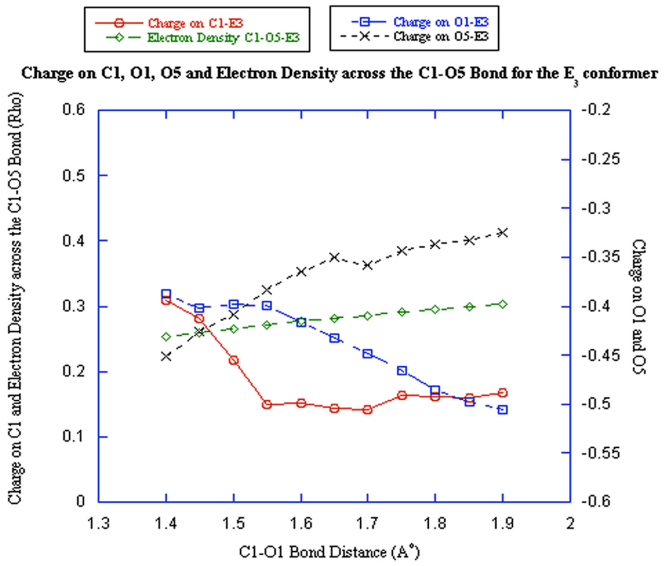


Figure B.8 The electron density across the C1-O5 bond, charge on C1, O1 and O5 as a function of the C1-OMe bond length for the E_3 conformer

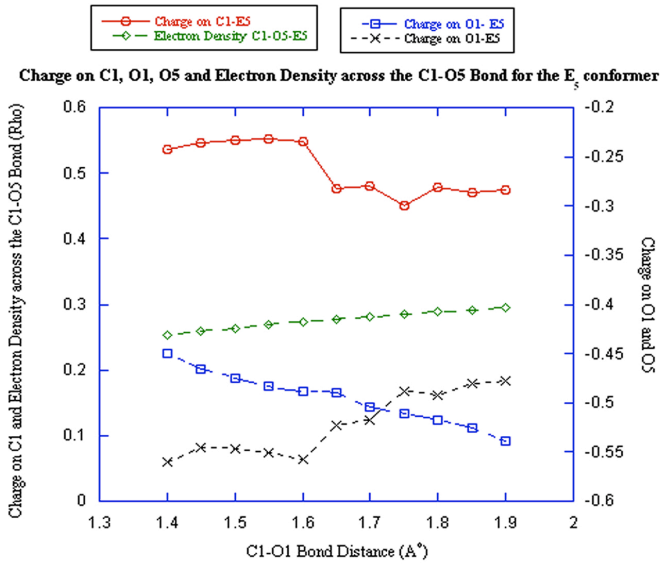


Figure B.9 The electron density across the C1-O5 bond, charge on C1, O1 and O5 as a function of the C1-OMe bond length for the E_5 conformer

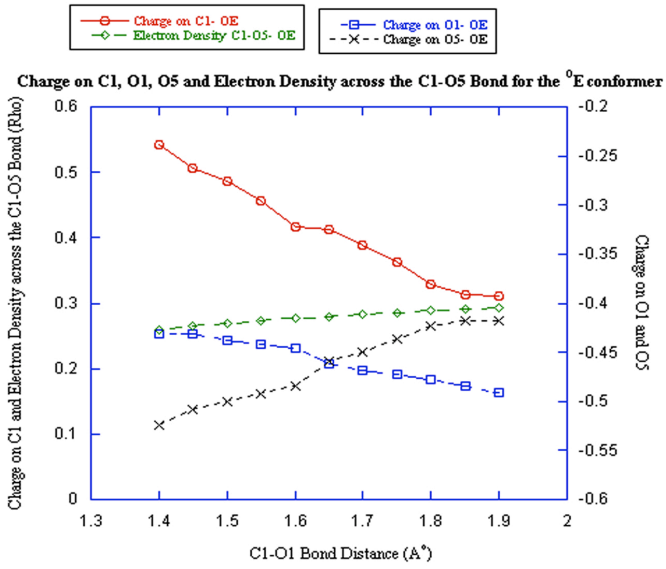


Figure B.10 The electron density across the C1-O5 bond, charge on C1, O1 and O5 as a function of the C1-OMe bond length for the ^{O}E conformer

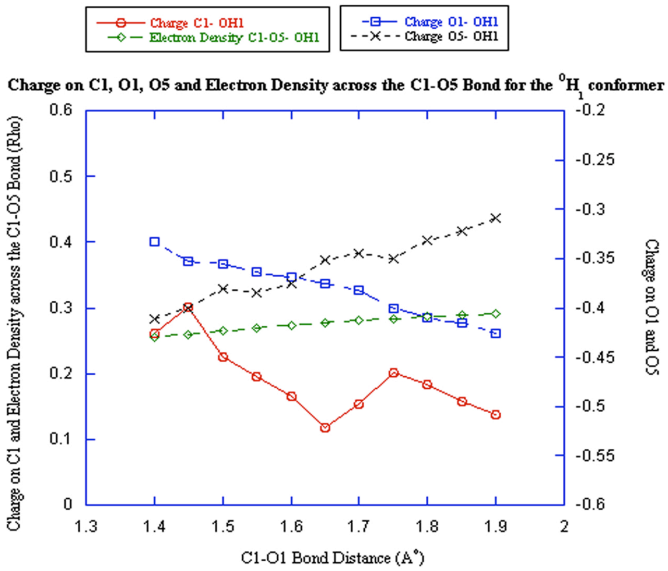


Figure B.11 The electron density across the C1-O5 bond, charge on C1, O1 and O5 as a function of the C1-OMe bond length for the $^{O}H_1$ conformer

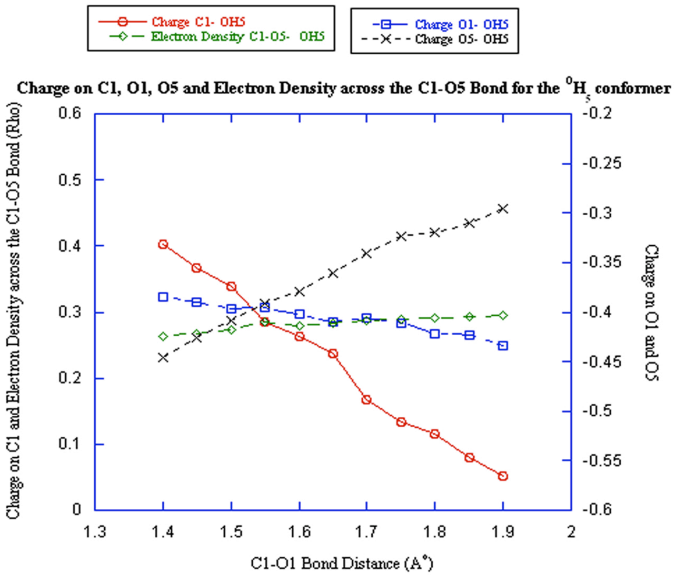


Figure B.12 The electron density across the C1-O5 bond, charge on C1, O1 and O5 as a function of the C1-OMe bond length for the $^{O}H_5$ conformer